DOCID: 3726612



National Security Agency
Fort George G. Meade, Maryland

Released by NSA on 16-Mar-2004 \n pursuant to the Freedom of Information Act\n Case #: 43698

ELEMENTS OF SIGNAL THEORY

National Cryptologic School
National Security Agency

PRINTED 1962
REPRINT 1963
REPRINT 1966

FOR OFFICIAL USE ONLY

DOCID: 3726612

ELEMENTS OF SIGNAL THEORY



A Textbook

Prepared For Use In

Basic Signal Analysis Training



National Cryptologic School
National Security Agency

PRINTED 1962
REPRINT 1963
REPRINT 1966

DOCID: 3726612

PREFACE

This text presents, in a simplified form, basic topics in signal theory which have been selected to form a foundation for a subsequent detailed study of the structure and parameters of signal systems. It has evolved from a series of monographs which have been used for some time in training non-Morse operator personnel in methods of identifying and analyzing non-Morse signal systems.

Although the monographs were prepared to support a specific course of instruction conducted by the Office of Training Services of the National Security Agency, requests from other organizations for complete sets of these necessitated the publication of the present text which brings them together under one cover. The style of the present text is such that it can be used for classroom instruction, self-study, and reference.

The information contained herein is available in numerous other documents. The chief merit of this presentation is that, while sophisticated mathematical explanations have been almost wholly eliminated, it is still directed to one who is a serious student of the subject. Each chapter contains a series of questions or problems to emphasize the key points which must be retained. Annexes follow the final chapter on signal theory (Chapter 7) and include a review of mathematics and an explanation of the operation of the cathode ray oscilloscope as it applies to pulse transmission analysis.

It is expected that in future editions of this text certain topics (notably single sideband) will be treated more fully.

DOCID: 3726612

ELEMENTS OF SIGNAL THEORY

TABLE OF CONTENTS

PART I: BASIC PRINCIPLES

Chapter 1	Waveform Analysis	3
	Characteristics of Sinewaves	4
	Waves in Three Dimensions	13
	Linear Mixing	16
	Fourier Components	28
	Bandwidth and Miscellaneous Complex Waveforms	31
Chapter 2	Pulse Analysis	43
	Pulses	43
	Pulse Nomenclature and Measurement	48
	Pulse Spectra and Bandwidth	56

PART II: MODULATION

Chapter 3	Amplitude Modulation	95
	Terms and Definitions	95
	Mathematical Possibility of AM	98
	Amplitude Modulated Waveform	103
	Amplitude Modulated Spectrum	109
	Bandwidth Calculations	116
Chapter 4	Single Sideband	135
	Terms and Definitions	135
	Mathematical Possibility of AM Single Sideband	137
	Single Sideband Waveforms	137
	The Single Sideband Spectrum	139
	Bandwidth Calculations	139

DOCID: 3726612

Chapter 5	Angle Modulation	143
	Terms and Definitions	143
	Mathematical Possibility of FM	148
	Frequency Modulated Waveform	149
	Frequency Modulated Spectrum	154
	Bandwidth Calculations	159
	Mathematical Possibility of PM	161
	Phase Modulated Waveform	162
	Phase Modulated Spectrum	166
	Bandwidth Calculations	168

PART III: MULTIPLEXING

Chapter 6	Frequency Division Multiplexing	175
	Information that is Frequency Multiplexed	176
	Baseband Synthesis (Frequency Division Multiplexing)	177
	The Baseband as a Modulating Wave	192
	Bandwidth Requirements	195
Chapter 7	Time Division Multiplexing	211
	Pulse Modulation	211
	Pulse Train Synthesis by Commutation (Time Division Mux)	223
	The Time Division Signal as a Modulating Wave	232
	Bandwidth Requirements	234

ANNEXES

Annex A	Mathematics for Modulation	245
	Common Fractions	245
	Decimal Fractions	255
	Positive and Negative Numbers	264
	Dimensional Analysis	273
Annex B	Oscilloscope Principles	285
	Cathode Ray Tubes	285
	The Basic Oscilloscope	297
	A Professional Laboratory Oscilloscope	302
	Operating Instructions	321

DOCID: 3726612

PART I

BASIC PRINCIPLES

Part I of this text, embracing Chapters 1 and 2, includes basic principles which must be thoroughly understood before the modulation theory treated in subsequent chapters can be studied and learned.

In an effort to simplify this series and to make it clearly understandable, mathematical explanations have been kept to an absolute minimum. By using Dimensional Analysis, very few mathematical formulas and equations will be required.

Waveform Analysis presents the basic concepts of signalling theory by a thorough discussion of the sinewave which is considered the fundamental building block of all other waveshapes. Linear Mixing will be discussed in detail to show the frequency components of complex waves.

In the discussion of Pulse Analysis will be found the theory of the complex waves so prevalent in communications today--squarewaves and rectangular waves--and the frequency spectrum occupied by each.

It is of utmost importance that every student put forth maximum effort in his study of this portion. In it lies the foundation for the succeeding chapters on modulation and multiplexing.

DOCID: 3726612

CHAPTER 1

WAVEFORM ANALYSIS

All transmitted intelligence is represented by electrical waves of various frequencies used singly or in combination. The manner in which these waves are treated in any particular situation may result in the generation of harmonics of a given wave; in a combination of waves of several frequencies in such a way as to generate, in addition to the parent components, new frequencies not harmonically related; or in altering the amplitude of the wave without changing its original form. In all but the last of these situations complex waveforms result. Obviously, these electrical waves are of prime importance in signal analysis. An extensive treatment of harmonic composition, bandwidth requirements, and equipment for handling these waves is the subject material of waveform analysis.

The basic waveform will be the sinewave which in turn will serve as the fundamental building block in the analysis of all other waveforms. A proper understanding of the sinewave is the foundation for the concepts of frequency, phase, modulation, complex waveforms, and bandwidth, and must therefore constitute a major portion of waveform analysis.

The objective of waveform analysis is to gain the requisite understanding of the sinewave characteristics; three dimensional presentation of waves; linear mixing of sinewaves of various frequencies, phase relationships, and amplitudes; the fourier components of complex waveforms; and bandwidth and miscellaneous complex waves.

Certain motions and/or changes repeat themselves at constant regular intervals. Such motions or changes are said to be periodic. In any periodic motion or change, the smallest identifiable portion of it is called a cycle, i.e., a cycle is the smallest identifiable part of a motion or change that is repeated after regular intervals of time.

Cycle is an important term in electrical communications and radio. Depending upon the desired outcome, the cycle may be represented graphically by either a circle or a sinecurve. The relationship between cycles and sinecurves (or sinewaves) will be expanded greatly in the pages that follow, but here a much briefer treatment of the relationship between a cycle and a circle will be given. This similarity between cycle and circle should be immediately obvious, therefore a definite distinction between the two must be made. A circle is a circuitous path in space, which if followed, would take you on a continuous round trip consisting of periodic journeys, each of which is an exact repetition of the previous

one. Each complete journey around the circle would constitute a cycle, that is, a complete and yet smallest identifiable part of a periodic motion or change.

Since the periodic motion takes some finite amount of time, cycle relates the repetitious journeys to time, whereas the circle relates this same journey to a path in space.

Similarities between circle and cycle make it convenient to use the terms of measurement which describe the circle to be used to describe the cycle also. This, in fact, helps to give a physical picture of the cycle. Examples of this sharing of terms may be found in expressions as, "...after so many degrees of the cycle...", or, "...a point in the cycle corresponding to $\frac{3\pi}{2}$ radians...."

This same cycle may be represented just as easily by a sinewave. In communications, cycles will almost always be represented by sinewaves and seldom by a circle. The standard example of representing cycles by circles in electronics will follow in this chapter where a circle inscribed by the point of a rotating vector is used to pinpoint values for plotting a sinewave.

I. CHARACTERISTICS OF SINEWAVES

Any signal that is a single pure frequency may be represented by a sinewave. If this signal were to fall within the bounds of the audio range, it would be recognizable as a pure audible tone, that is, it would not sound like a note on a piano or a trumpet but would be as the note from a signal generator. In fact, the idea of frequency and sinewave are so closely related that in electronics the term frequency can only properly apply to a sinewave.

Sinewaves should be understood correctly. A true sinewave satisfies the formula (which shall be explained shortly), $a = A \sin \theta$, where θ (the Greek letter theta) is used to designate some angle. A wave that meets this requirement can have only one possible shape--a sinusoidal shape. Any wave that is nearly a sinewave or almost satisfies the formula just given does not represent a pure tone or a single frequency but indicates distortion which is the presence of additional frequencies. The main point of emphasis here is that not just any rounded off wave can be considered a sinewave, and therefore, care should be exercised in drawing sinusoidal waveforms.

DOCID: 3726612

A. Sines

The student may remember from previous schooling the terms sine, cosine, and tangent. It is not within the scope of this course to go into an explanation of these functions except that these terms refer to certain ratio-type relationships in a right triangle. In this course the interest is in the sine-function of angles, and it must suffice to say that every angle has a respective sine. Below are listed certain angles from 0° to 360° with their respective sines:

TABLE OF SINES

$\sin 0^\circ = 0$
$\sin 30^\circ = .5$
$\sin 60^\circ = .866$
$\sin 90^\circ = 1.$
$\sin 120^\circ = .866$
$\sin 150^\circ = .5$
$\sin 180^\circ = 0$
$\sin 210^\circ = -.5$
$\sin 240^\circ = -.866$
$\sin 270^\circ = -1$
$\sin 300^\circ = -.866$
$\sin 330^\circ = -.5$
$\sin 360^\circ = 0$

It should be noticed that these values increase from 0 to 1 and decrease from 1 to 0 for angles from 0° to 180°. The same sequence of values exists for angles between 180° and 360° except that now the sines are preceded by a minus (-) sign. Two points that should be remembered are: (1) any circle or cycle has 360°; and (2) sine is frequently abbreviated sin.

B. Generating the Wave

Assume a vector (an arrow representing magnitude and direction) of a given length, A, lying in a horizontal plane and pointing to the right. Now, draw a line of some length a short distance to the right of the vector and in a direct line with it. This line will be a zero reference around which a sinewave will be drawn. To be useful this line must be divided into increments representing time. Because the total length of this line and any part of this line represents time, it is customary to refer to it as a time-base. (See Figure 1)

DOCID: 3726612

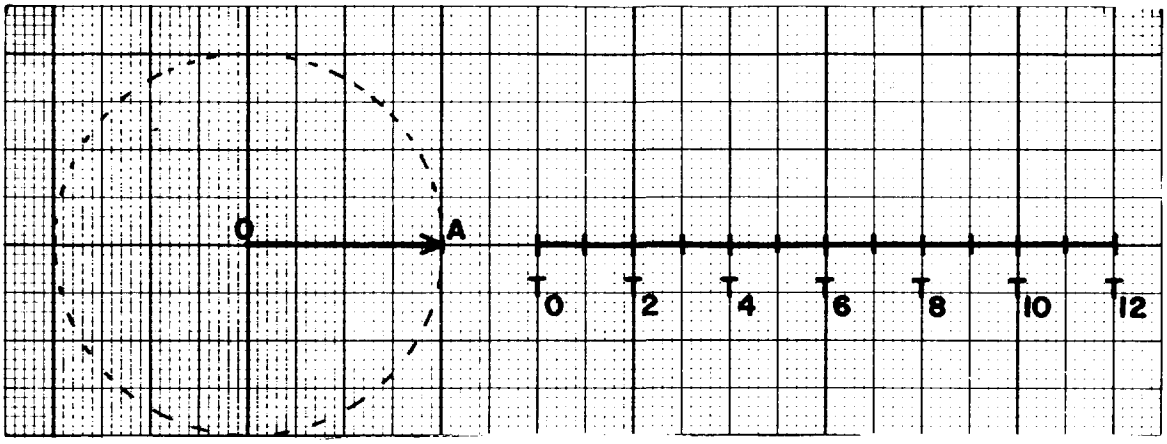


Figure 1 - A Vector and a Time-Base

If the vector, A, rotates in a counterclockwise direction and at a uniform rate, it will take a certain amount of time for the vector to move a given position. Assume that it takes the time from T_0 to T_1 for the vector to rotate from its 0°, or at rest, position to the 30° position. Plot a point above T_1 in a direct line across from the point of the vector. If it took such a length of time for the vector to rotate 30°, it will take an identical period of time to move an additional 30°, that is, rotating the vector to the 60° position takes it to time T_2. Now plot a point above T_2 but directly across from the point of the vector as before. Repeat the procedure for every 30° for the full rotation of the vector through one complete cycle of 360°. When all points have been plotted, draw a continuous line from T_0 through each of the plotted points to T_{12}. This continuous curved line, if carefully drawn, is the shape of a true sinewave. (See Figure 2)

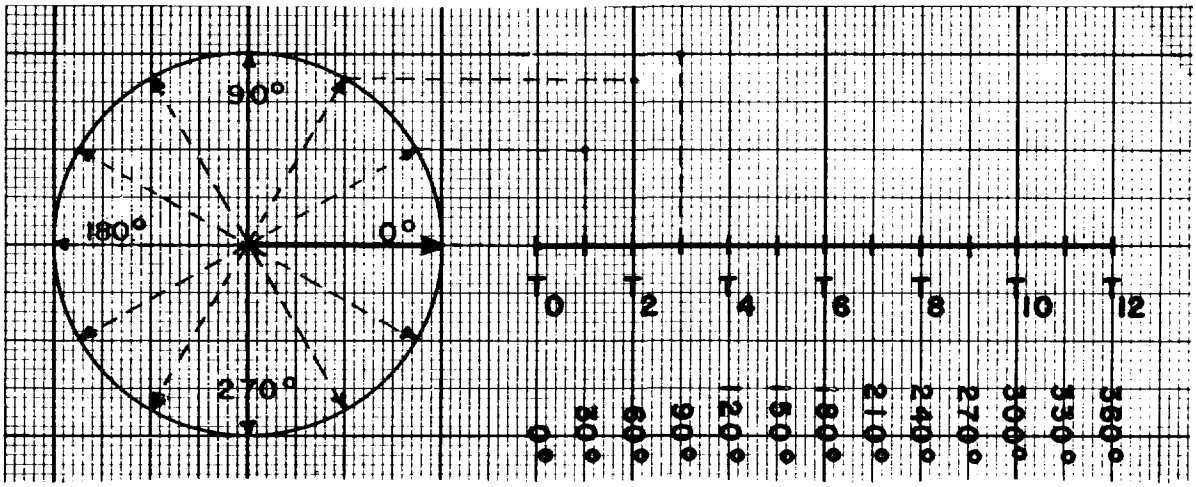


Figure 2 - A Rotating Vector Generating a Sinewave Pattern

Observing the sinewave just drawn several things should be noted: (1) Rotating the vector through 360°, or one cycle, takes time. Therefore the length of the reference line or time base represents the time required for the cycle. In the illustration just used, each time increment represents the time required for 30° of rotation. (2) At any point in time along the reference line, the sinewave has some definite amplitude. This amplitude is called the instantaneous amplitude, and it is represented by a in the equation $a = A \sin\theta$. (3) There is a certain maximum amplitude, A in the formula, which occurs once above the reference line and once below it for each cycle of the wave at points in time corresponding to 90° and 270° of rotation. (4) The maximum amplitude, A, is identical with the length of the vector. (5) The sinewave is a function of an angle that varies uniformly with time.

C. A Sinewave Using the Formula

It is also possible to repeat the results obtained from the rotating vector by using the table of sines provided in this chapter and the formula, $a = A \sin\theta$. Observe that the term, sin, in the formula refers to the sine of any specified angle and that θ represents any angle, which in this case is any specific angle for which the table of sines indicates a value.

D. How to Proceed

The equation, $a = A \sin\theta$, means that the instantaneous amplitude, a, is equal to the maximum amplitude, A, multiplied by the sine of the specified angle. For example: let A = 10 volts. If at 0 degrees the instantaneous amplitude, a, is equal to 10 times the sine of 0° (i.e., $a = 10 \sin 0°$) and from the table of sines, the sine of 0° = 0, then the value of a at 0° is 0 because $a = 10 \times 0$.

In a similar manner, the equation can be solved at 30° after looking up the sine of 30° in the table. Here again $a = A \sin\theta$ or $a = 10 \times .5$ or $a = 5$. Each of these points should be plotted on the graph provided (Figure 3). After completing this sinewave, plot two more sinewaves in the same block: one having a peak amplitude of 20 volts, and one having a peak amplitude of 30 volts.

DOCID: 3726612

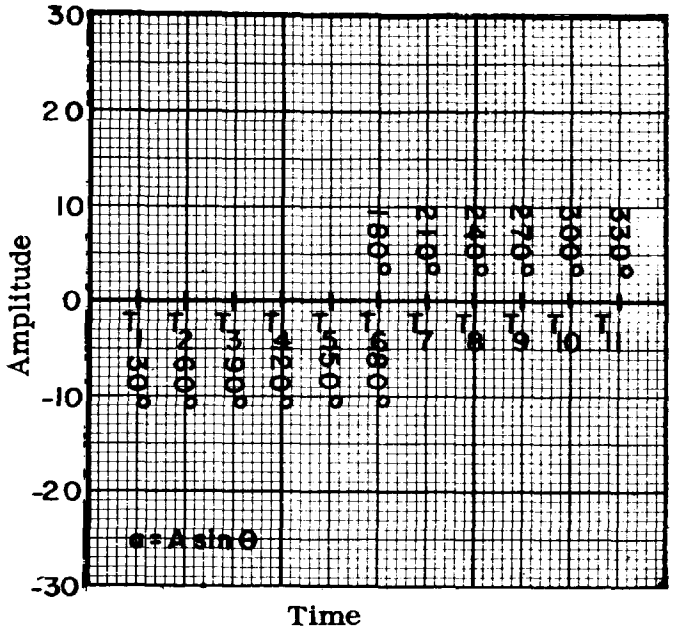


Figure 3

Even though the three waves differ in amplitude, they still satisfy the equation, $a = A \sin \theta$. In other words, the instantaneous amplitude of any sine-wave is directly proportional to the sine of the angle that exists at that particular instant of time.

E. Radians

Due to the similarity between cycles and the circles often used to represent them, convenience dictates that certain measurements used to describe the one will be common to both. These measurements are degrees and radians. It is, then, but one step more to apply either degrees or radians to the sinewave. In fact, it is frequently more convenient to measure angles in terms of radians instead of degrees. The radian can be described as follows: Divide the circumference (distance around) of a circle--any circle--by the diameter (distance through) of that circle and the result will be 3.141592.... This value is called pi and is symbolized by the Greek letter π. Therefore, the circumference is π times larger than the diameter, or $C = \pi D$. Using the radius (one-half the diameter) instead of diameter, the formula becomes $C = 2 \pi r$.

Any type of ruler can be used to measure the circumference of the circle. It could be measured in inches, centimeters, picas, so many lengths of string, or

DOCID: 3726612

even a unit equal to the length of the radius of the given circle known as a radian. Since the length of the radian would be equal to the length of the radius, and the circumference equals 2 π times the radius, the circumference would also be 2 π times as long as the radian. Note well that the radian is an actual measure and is related to the circumference by the factor 2π.

Any circle may be divided into 360°. These degrees may be marked along the circumference. The circumference can, then, be laid off in either of two units commonly used for dividing it into segments--either degrees or radians.

Since $C = 2\pi r$ and the radius is equal to the radian, then $C = 2\pi$ radians. Thus, the conclusion is reached that $360° = 2\pi$ radians. Dividing both sides of this equation by 2π the result will be that one radian equals 57.2957914...degrees. This figure is usually rounded off to 57.3°.

The radian is a very useful measure and is used many times in radio and electronics. It can be used in place of degrees as a measure of angles. Frequently radian measure is called π measure because of its relationship to π. Figure 4 illustrates a sinewave divided into radians.

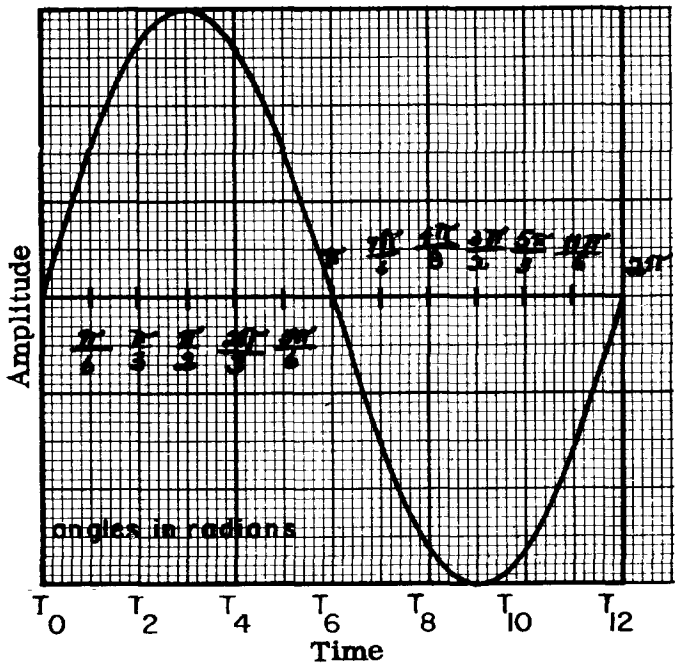


Figure 4

To convert degrees to radians in terms of π, multiply the given number of degrees by $\frac{2\pi}{360}$. Example: $30^{\circ} \times \frac{2\pi}{360} = \frac{60\pi}{360} = \frac{\pi}{6}$ radians. The specific value may be obtained by evaluating π and performing the indicated division and/or multiplication. To convert radians to degrees multiply the given number of radians by $\frac{360^{\circ}}{2\pi}$. Example: $\frac{\pi}{6}$ radians $\times \frac{360^{\circ}}{2\pi} = \frac{360^{\circ}}{12} = 30^{\circ}$. If the radians are expressed as a specific number, π must be evaluated.

The important thing to remember about the radian in this course is not so much how it is derived, but rather that it represents 57.3°, and that there are 2π radians in a complete cycle.

F. Period

A definite period of time is required for the full cycle of 360°. This period of time is called the period of the wave. If the time for the 360° or 2π radians is one second, then the period of the sinewave is one second. Since only one cycle has occurred in this time, then the frequency of the wave is one cycle per second. It should be evident that the period and the frequency are closely related. In fact, by closer inspection it will be obvious that period and frequency are just two different ways of saying the same thing. The frequency is measured in cycles per second (cycles/second), whereas the period is measured in seconds per cycle (seconds/cycle). Both of these terms are useful, for, if one quantity is known the other one can be found. For example, frequency can sometimes be measured very easily by means of Lissajous patterns which will be discussed later. Then it becomes a simple matter to calculate the period of the wave. On the other hand, there are times when it is desirable to determine the lowest frequency of a complex wave--a measurement not as easily made. Here it is simpler to measure the period of the wave and then calculate the frequency.

G. Period-Frequency Conversion

The matter of converting period into terms of frequency and frequency into period is an easily followed procedure.

1. Call the known quantity by its full proper name (i.e., call frequency "cycles per one second" instead of just "cycles," and call period "seconds per one cycle" instead of just "seconds").

2. Express this known quantity mathematically as $\frac{60 \text{ cycles}}{1 \text{ second}}$ or $\frac{1 \text{ second}}{60 \text{ cycles}}$.

3. Invert the mathematical expression to obtain the desired unknown quantity. (This inverted form is called a reciprocal. Therefore, frequency is the reciprocal of the period, and the period is the reciprocal of the frequency.)

4. Perform whatever division is necessary to obtain the desired answer in its simplest and best form. (Notice how the above steps are employed in the following examples.)

EXAMPLE 1: Assume a frequency of 1000 cycles. What is the period?

Step 1: 1000 cycles should be called "1000 cycles per one second."

Step 2: Expressed mathematically this is $\frac{1000 \text{ cycles}}{1 \text{ second}}$.

Step 3: The period is the reciprocal (inverted form) of the frequency.

$$f = \frac{1000 \text{ cycles}}{1 \text{ second}} \quad \text{(invert)} \quad \frac{1 \text{ second}}{1000 \text{ cycles}} = P$$

Step 4: Perform the indicated division.

$$P = \frac{1 \text{ second}}{1000 \text{ cycles}} = \frac{.001 \text{ seconds}}{1 \text{ cycle}}$$

(Call this 1 millisecond per cycle)

EXAMPLE 2: Assume a period of .5 milliseconds, what is the frequency?

Step 1: .5 milliseconds should be called ".5 milliseconds per one cycle."

Step 2: Expressed mathematically this is $\frac{.0005 \text{ seconds}}{1 \text{ cycle}}$.

Step 3: The frequency is the reciprocal (inverted form) of the period.

$$P = \frac{.0005 \text{ seconds}}{1 \text{ cycle}} \quad \text{(invert)} \quad \frac{1 \text{ cycle}}{.0005 \text{ seconds}} = f$$

Step 4: Perform the indicated division.

$$f = \frac{1 \text{ cycle}}{.0005 \text{ seconds}} = \frac{2000 \text{ cycles}}{1 \text{ second}}$$

DOCID: 3726612

H. Phase

In addition to period and frequency there is another time relationship of importance. If it were desired to indicate that a certain part of a cycle occurs at a particular instant of time, the terms frequency and period would be awkward. However, this particular relationship could best be expressed by the term phase.

Phase is a measure of time in terms of degrees or radians. If the sinewave is assumed to begin at zero amplitude, then at a phase angle $\frac{\pi}{2}$ radians or 90° later in time the peak amplitude will have been reached. Similarly any other point on the wave may be defined as occurring at some phase angle expressed in radians or degrees in time.

I. Frequency as Rate of Change of Phase

Viewing the sinewave along the time base, it is obvious that the phase angle is continuously changing with time. It can even be said that after so many degrees or radians of phase change a certain span of time will have elapsed. If 360° or 2 π radians of phase change occurs the elapsed time will be equal to the period of the sinewave. But 2 π radians or 360° is one cycle so what has actually been found is a cycle and the time required for it. This is identical to the measurement of frequency. The conclusion is that frequency is the time rate of change of phase.

J. Harmonics

The terms harmonic and sinusoidal are used interchangeably and describe a sinewave type of variation, hence any pure frequency is a sinewave and a simple harmonic variation. In electronics the term harmonic usually designates a whole number multiple of some basic sinewave.

If, during a specified period of time, one cycle of a sinewave is completed and then two cycles of another sinewave occur in the same span of time, the original wave would be called the first harmonic and the other wave the second harmonic. The frequency of the original wave is called the fundamental frequency. The frequency of the other wave is obviously twice as great as the fundamental frequency since there are twice as many cycles of it in the designated interval of time. This is why it is called the second harmonic. If a multiple three times the frequency of the fundamental or first harmonic wave were also produced it would be named the third harmonic. Therefore, the name of a particular harmonic is,

DOCID: 3726612

then, designated by the number of times its frequency is greater than the fundamental. (Unfortunately, the identification of a particular harmonic in music is not the same as that used in physics or electronics.)

A wave may be distorted to produce harmonics but under no condition will it generate sub-multiples. Therefore, the term subharmonic, although it may be encountered on occasions, is more a term of convenience than of fact and should be avoided.

K. Summary of the Sinewave

The plot of a sinewave is a graph of the sine of an angle that uniformly increases with time through 360°. This wave may be represented by an equation or formula namely, $a = A \sin \theta$. From the formula or by graphical analysis, it can be proven that a sinewave has a peculiar shape which distinguishes it from all other wave shapes. For this reason semicircles should not be used to represent a sinewave.

A sinewave can be completely defined by noting its frequency (or period), peak amplitude and phase.

II. WAVES IN THREE DIMENSIONS

Of several methods of viewing a wave (not necessarily a sinewave) three that are of analytical importance are:

A. Amplitude as a Function of Time

This method illustrates the actual shape of the wave and is the method employed here in drawing sinewaves. It is this type of presentation that is obtained on an oscilloscope (see Figure 5).

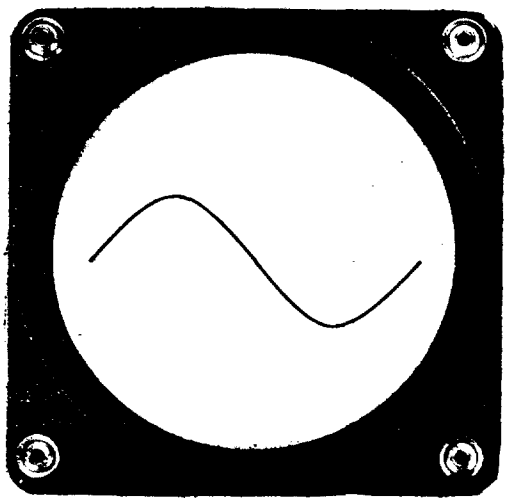


Figure 5

DOCID: 3726612

B. Frequency as a Function of Time

This method illustrates the time duration of the sinewave components of some complex wave. The presentation tells nothing of the shape of the components (although any pure tone is always a sinewave) but it does indicate what the frequency is and how long it lasts (See Figure 6A). The indication that a frequency is present is simply a straight line. This straight line indicates the presence of a sinewave. Where several frequencies are present, they will appear as parallel horizontal lines separated vertically along the frequency axis to a position corresponding to the frequency of the individual sinewaves present (see Figure 6B). Laboratory equipment typical of giving this sort of presentation is the Sonograph.

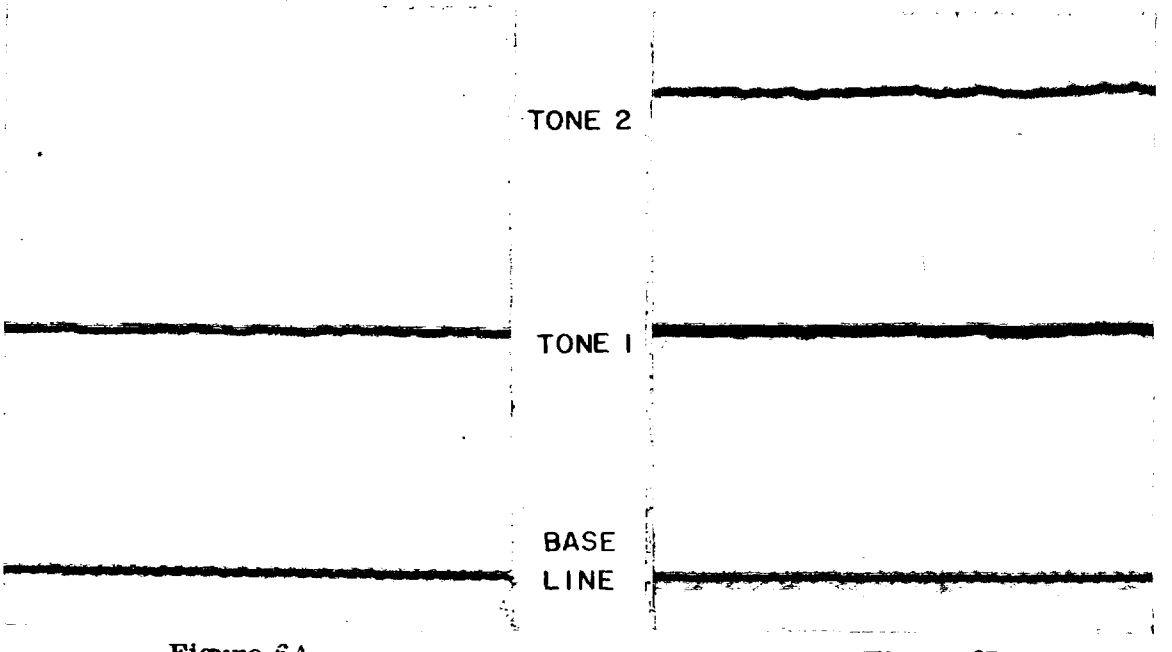


Figure 6A

Figure 6B

C. Amplitude as a Function of Frequency

The third method of considering a wave represents the wave as a straight vertical line whose length is directly proportional to the amplitude of the sinewave, and whose position along a horizontal base line is an indication of the frequency of the wave (see Figure 7).

DOCID: 3726612

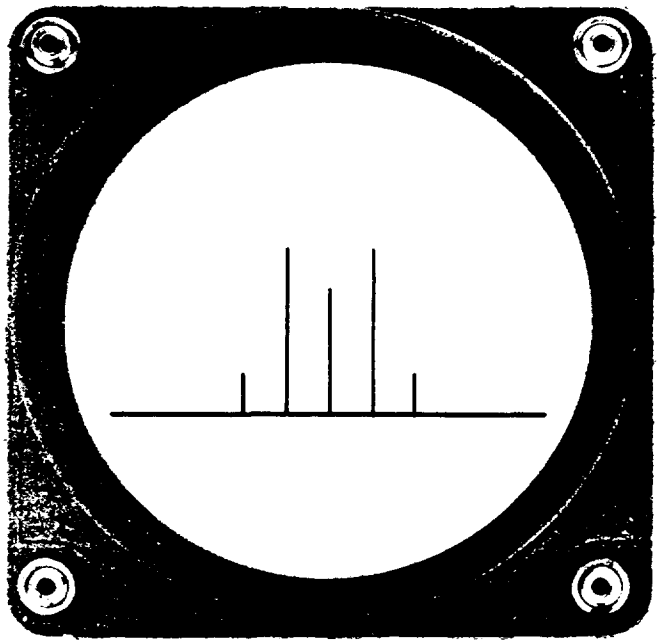


Figure 7 - Typical Spectrum Analyzer Presentation

Where sinewaves of several different frequencies and amplitudes are present, they will appear as parallel vertical lines whose individual lengths indicate their relative amplitudes, and whose separation and position along the horizontal frequency base corresponds to the frequencies of the individual sinewaves present. Equipment capable of this presentation is the spectrum analyzer.

A composite picture of these three presentations may be clearly fixed in mind by looking at the block in Figure 8. The vertical dimension (height) represents amplitude, the horizontal dimension (length) represents time and the remaining dimension (width) represents frequency. From the equipment point of view, the front face of this block is the amplitude versus time relationship as presented on the oscilloscope, the top face is the frequency versus time relationship as seen on the Sonograph and the side face is the amplitude versus frequency relationship shown on a spectrum analyzer. The figure shows the patterns as they would appear on these three instruments for a single sinewave.

DOCID: 3726612

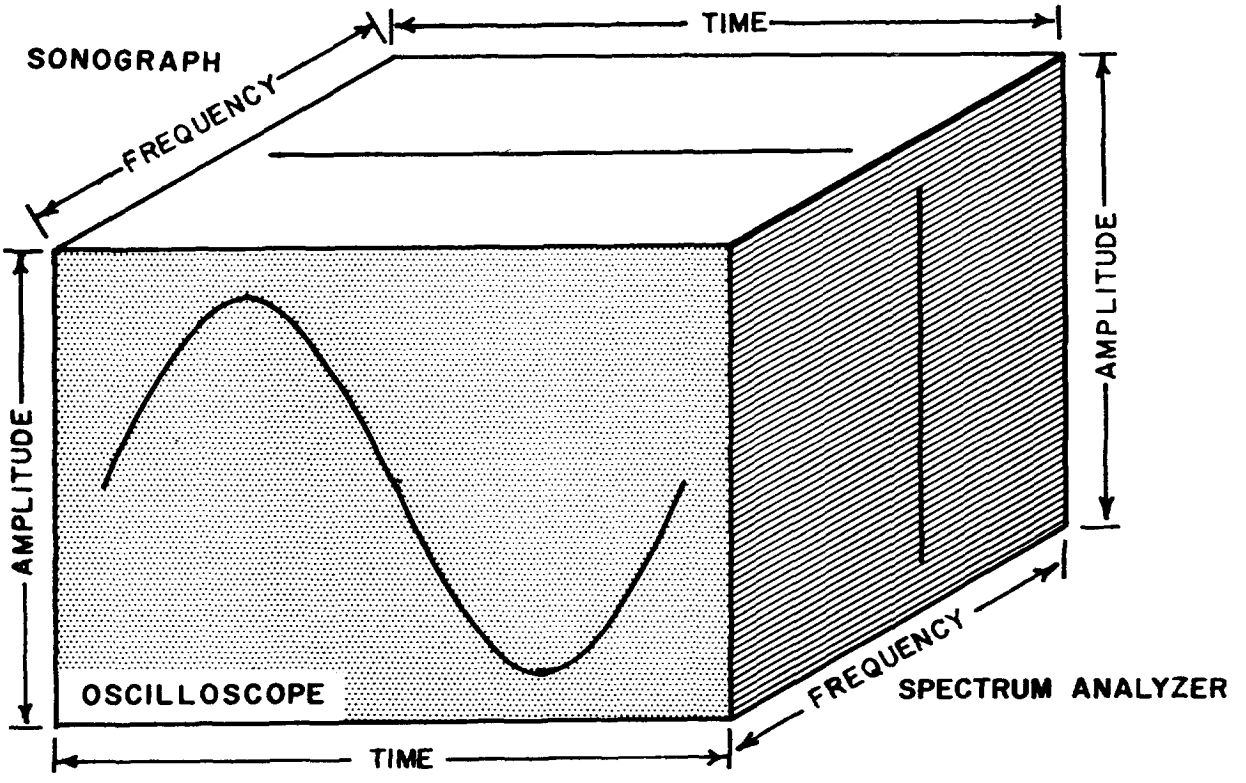


Figure 8 Three Dimensional Presentation

III. LINEAR MIXING

Sinewaves of various frequencies, phases, and amplitudes may be combined to produce some complex waveform. This mixing may be either of two types: linear or non-linear. Linear mixing is also called adding. An elaboration on it and its effects will occupy the balance of this lesson. Non-linear mixing is usually called heterodyning and will constitute a large part of the study of modulation.

A. Linearity

Perhaps it would be well to describe what is meant by the terms linear and linearity. If a car, driven on the open highway, gives 22 miles to the gallon, then for each gallon of gas burned in the open highway driving, the distance to the destination would be decreased by 22 miles. If a graph were made of the relationship of miles as a function of gallons, the result would be a straight line. In other words, the relationship of miles to gallons is linear (see Figure 9).

DOCID: 3726612

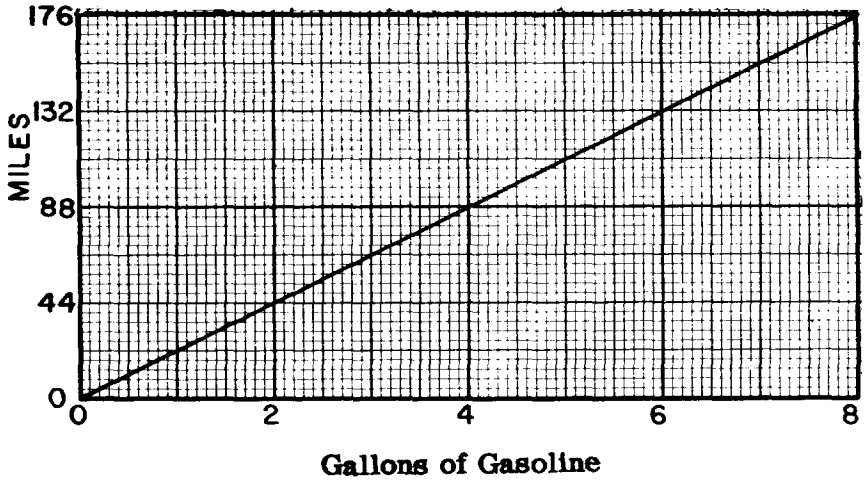


Figure 9

Therefore, the prediction can be made as to how far the car can go with a specified number of gallons of gas. The prediction will be fairly accurate because with a linear (or straight line) relationship, effect is always directly proportional to cause.

On the other hand, if the same car were driven in the city, the mileage per gallon would be considerably different. The mileage would be influenced by the number of starts, stops, waits, and shifts of the gears. It would be impossible to plot a straight line (linear) relationship of miles to gallons for this type of driving. The relationship is now non-linear. The non-linearity is due to the fact that effect (mileage in this case) is no longer directly proportional to the cause (gallons of gas).

It can be concluded that whenever effect is directly proportional to cause, linearity occurs and whenever effect is not directly proportional to cause, non-linearity occurs.

B. Characteristics of Linear Mixing

Two waveforms may be mixed and the mixture will result in a complex waveform. If the device in which they are combined is linear, the two waves, which may be considered as causes, will each produce its own individual effect in the output of the device. The resultant output will be the sum of the individual effects of the individual causes. If, on the other hand, the device in which the two

DOCID: 3726612

waves were mixed were a non-linear device, the device would add to the effects of the two waves a few effects of its own. The overall effect in the output of the device would be a complex waveform consisting of the sum of the effects of the two original signals (the intended causes) and any side effects introduced by the mixing device.

C. Examples of Linear Mixing

1. Sinewaves of Like Frequency, Amplitude and Phase.

Consider two sinewaves of the same frequency, amplitude and phase as in Figure 10. Adding these two waves point by point, notice that when wave A

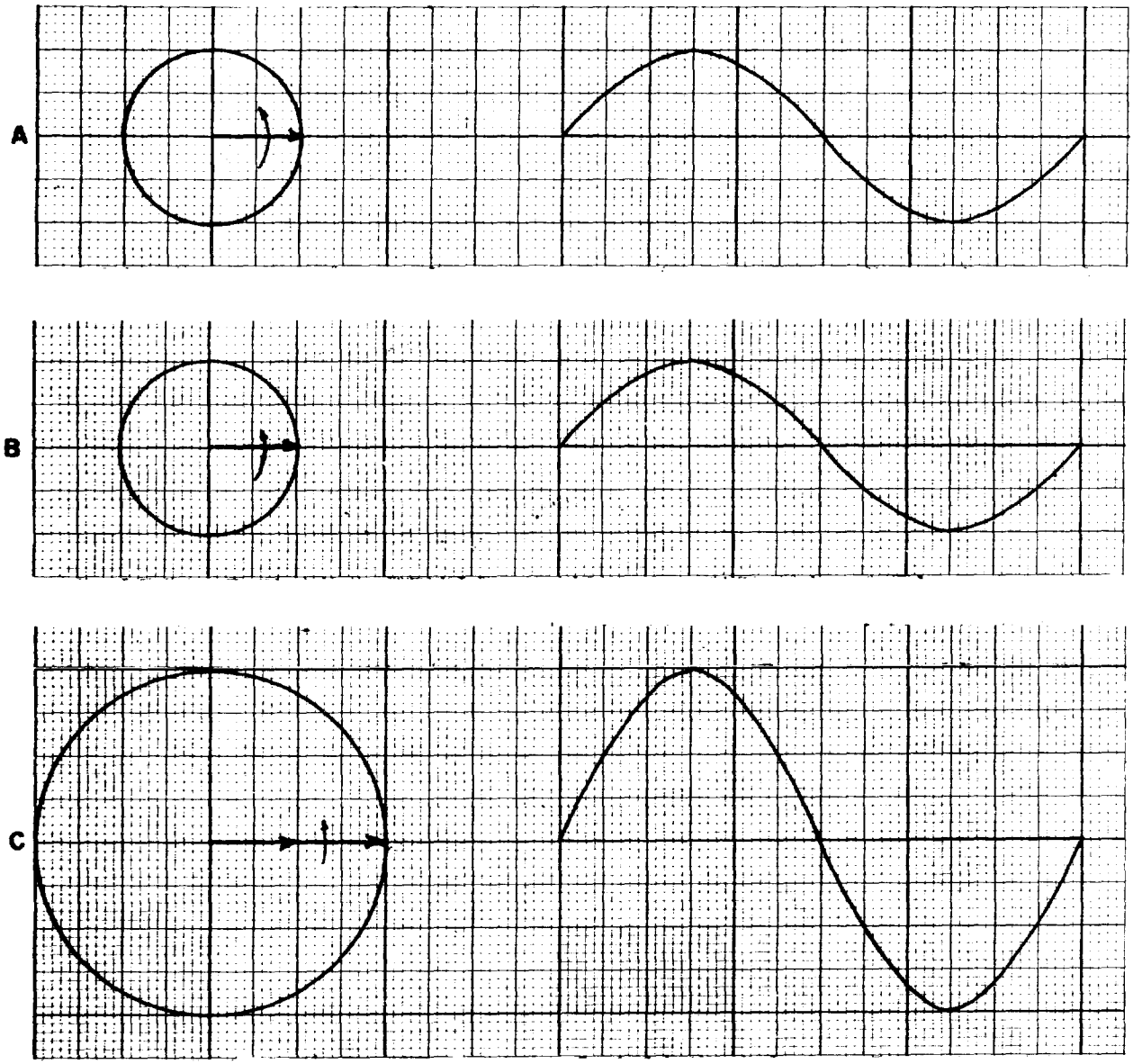


Figure 10 - Like Frequency, Phase, and Amplitude

goes one volt positive, wave B also goes one volt positive and the total voltage applied at this point to form the new wave C is two volts. Earlier, a sinewave was generated by means of a rotating vector. Here are two sinewaves, each of which could be generated by a rotating vector. Since these two vectors, A and B, are rotating in the same direction and point the same way at the same time, their lengths may be added directly to give a resultant vector, C.

2. Sinewaves of Like Frequency and Amplitude but 90° Phase Difference.

Now take two vectors and their waves representing identical frequencies, and identical amplitudes but 90° different in phase. As before, at any instant of time the amplitude of the resultant wave is the sum of the instantaneous amplitudes of the two waves, A and B. From the standpoint of vectors, it is clear that the vectors of the two waves are at right angles to each other. The resultant of these two vectors may be found by adding a phantom line parallel to each of the vectors and making contact with the point of each. A resultant vector may then be drawn from the origin of the two parent vectors to the corner of the parallogram or square opposite the origin. Observe that the wave generated by the rotation of this resultant vector is the wave resulting from the algebraic addition of the two waves, A and B (See Figure 11).

3. Sinewaves of Like Frequency and Amplitude but 180° Phase Difference.

To illustrate further, suppose there are two waves to be linearly mixed which have identical frequency and amplitude but differ in phase by 180°. Obviously, if the waves are 180° out of phase, their vectors will also be 180° out of phase. This means that the vectors will be equal and opposite at every point of their rotation and will cancel each other at all times. By reason of this cancellation no vector is shown for the resultant wave (see Figure 12). It will be left to the student to determine what the resultant wave will look like.

4. Sinewaves of Unlike Frequency and Amplitude, but Initially in Phase.

Figure 13 is a case of two waves added together that have a large frequency difference and some difference in amplitude. It should be noted that the higher frequency wave appears to ride on the lower frequency wave. Also of note is the fact that unlike the three previous waveforms, these two parent waves add together to form a complex wave. The higher frequency parent tends

DOCID: 3726612

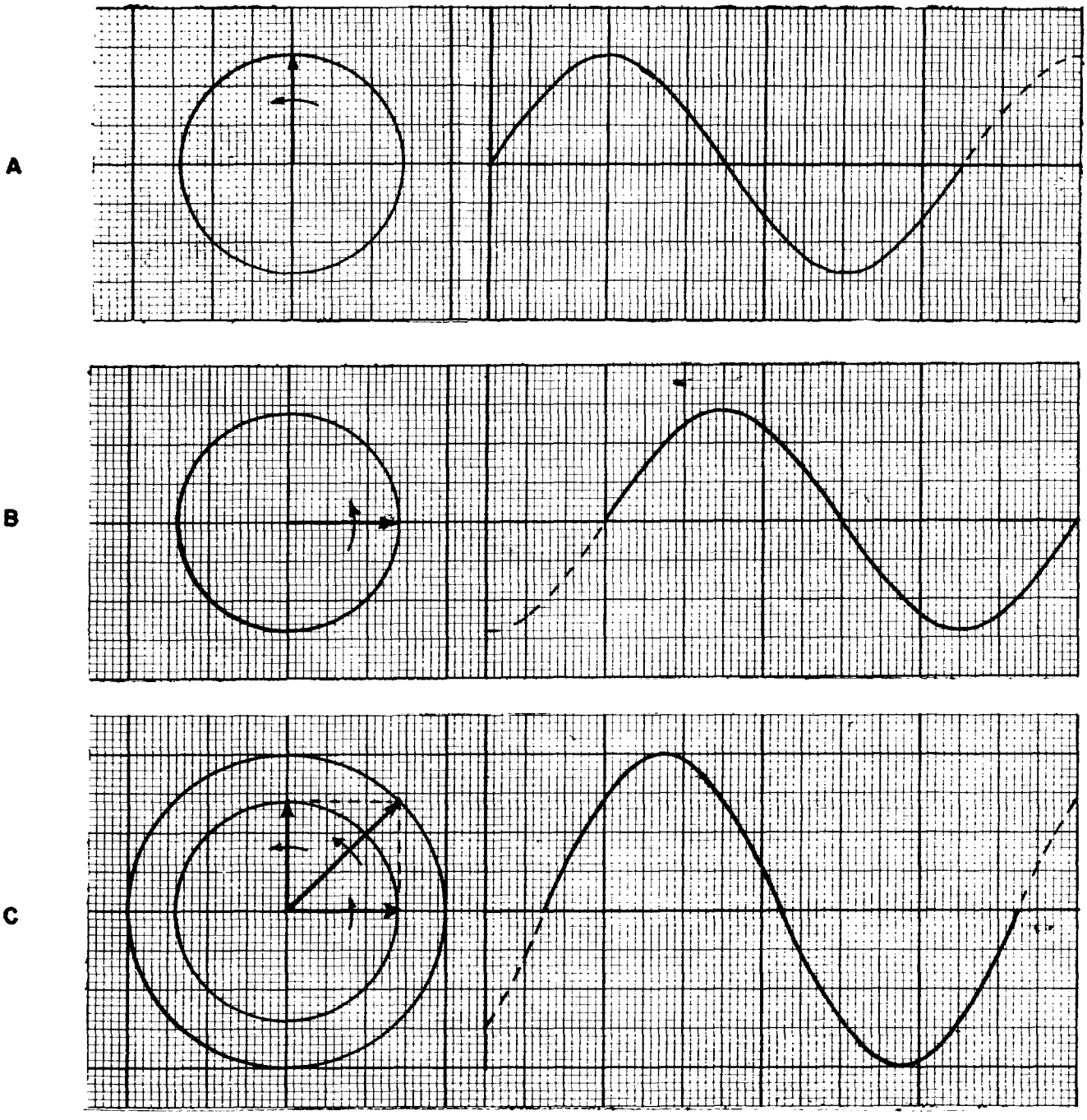


Figure 11 - Like Frequency, Like Amplitude, but Unlike Phase (90°)

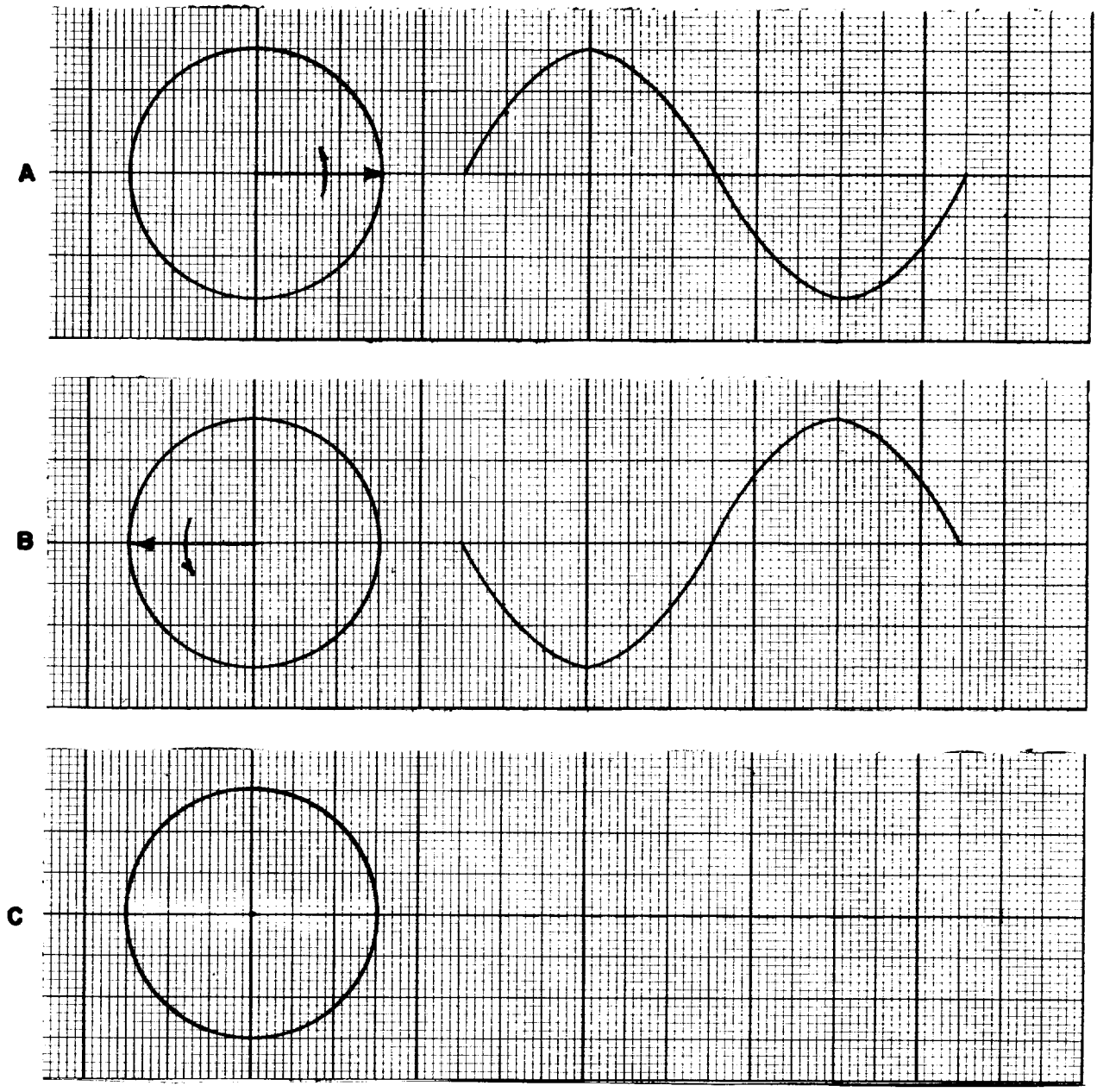


Figure 12 - Like Frequency, Like Amplitude, but 180° Phase Difference

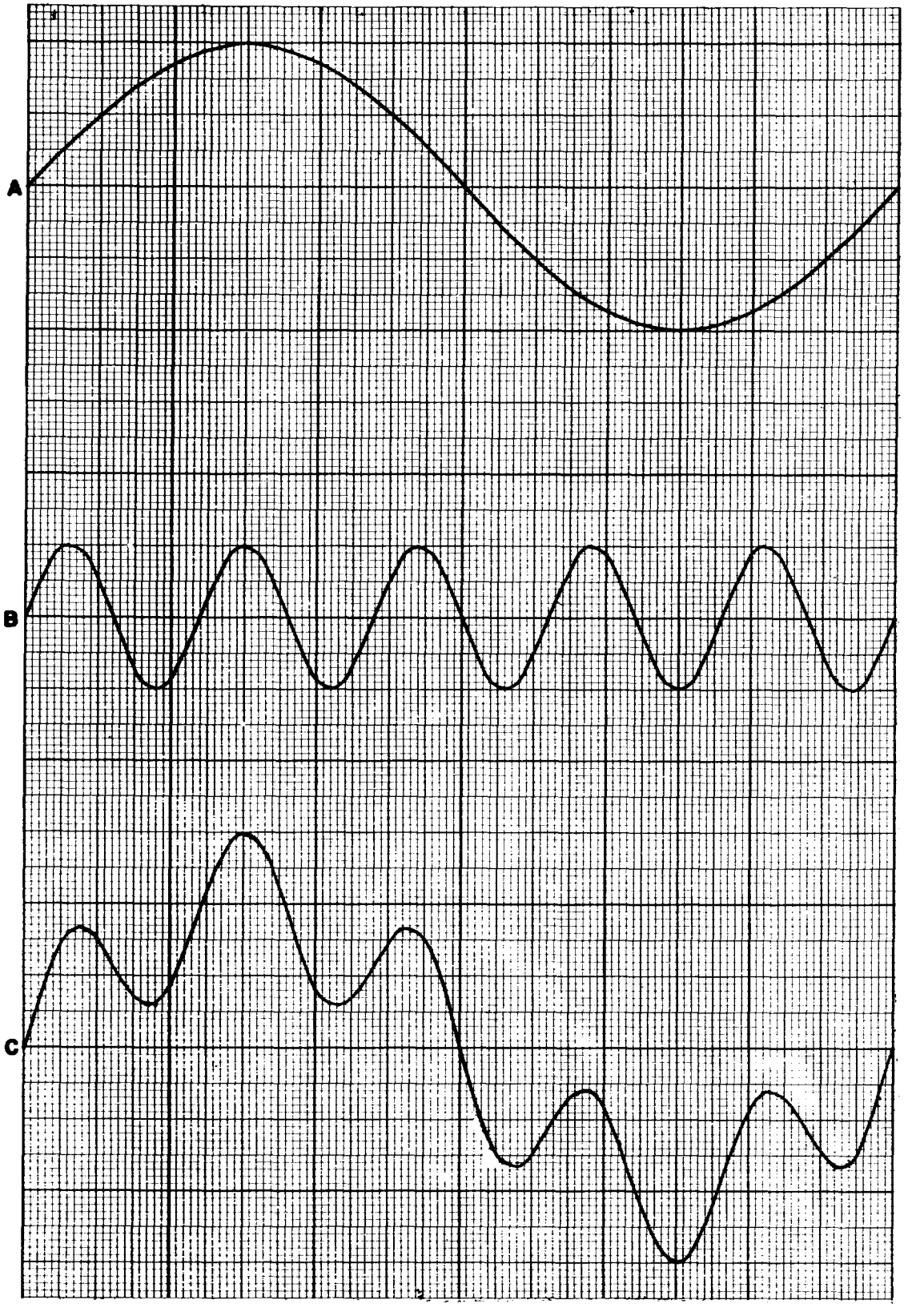


Figure 13 - Sinewaves of Unlike Frequency and Amplitude but Initially in Phase

DOCID: 3726612

to make the sides of the resultant complex wave more steep than the sides of the low frequency parent. A point to remember is that a steep rise or steep decay of a wave always indicates the presence of high frequency components. The reason for this is that a sharp rise or fall evidences a rapid change per unit interval of time.

5. Cosinewave Added to a Squarewave.

The squarewave itself is composed of sinewave components but for the moment consider it as a unit in itself. The effects of the addition of waveforms may be further emphasized by the superposition of a squarewave and a cosinewave. From Figure 14 it can be seen that the two waves are displaced in such a manner that they do not change polarity at the same time. (They cannot be related as to phase since only a sinewave or cosinewave can have phase.) When adding these two waves point by point notice that the cosinewave which has been varying about a zero dc level now varies about the changing level provided by the flat top and bottom of the squarewave. Of course, the result of the linear mixing or adding is that the complex wave is neither a cosinewave nor a squarewave.

6. Squarewave from Sinewaves.

If two sinewaves are added, one of which is three times the frequency (third harmonic) and one-third the amplitude of the other wave, and added in such a manner as to have the two waves initially in phase, two things occur. The sides of the resultant wave will be more nearly vertical than the sides of either of the parent waves, and the top and bottom of the resultant wave, though irregular, will be flatter than the parent waves (Figure 15A). If a sinewave five times the frequency (fifth harmonic) and one-fifth the amplitude of the first wave is added to the wave already obtained and assuming that all three of the sinewaves are initially in phase, the sides of the new resultant wave will be steeper than before, and though irregularities will still be present, the top and bottom will become still flatter (Figure 15B). The addition of more odd harmonics initially in phase with the first wave and decreasing in amplitude as the reciprocal of the harmonic order, will continue to flatten the top and bottom of the resultant and steepen its sides in such a manner as to make it become more and more nearly square. Theoretically, if an infinite number of odd multiples of the first wave having the proper phase and amplitude relationship were added together, a perfect squarewave would be the end result.

The generation of a squarewave from sinewaves is dependent upon waves adding algebraically so as to reinforce or cancel either fully or partially

DOCID: 3726612

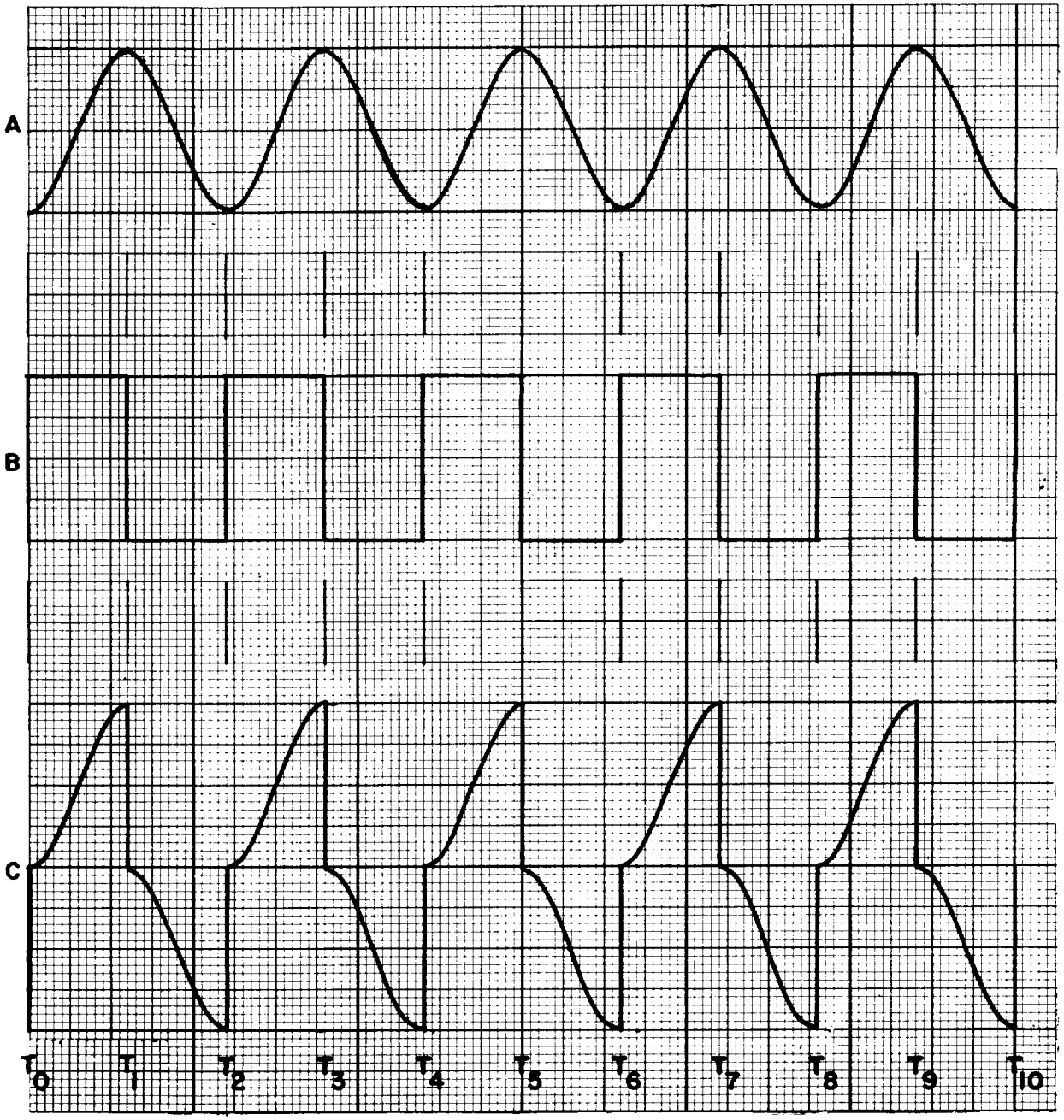


Figure 14 - A Cosinewave Added to a Squarewave

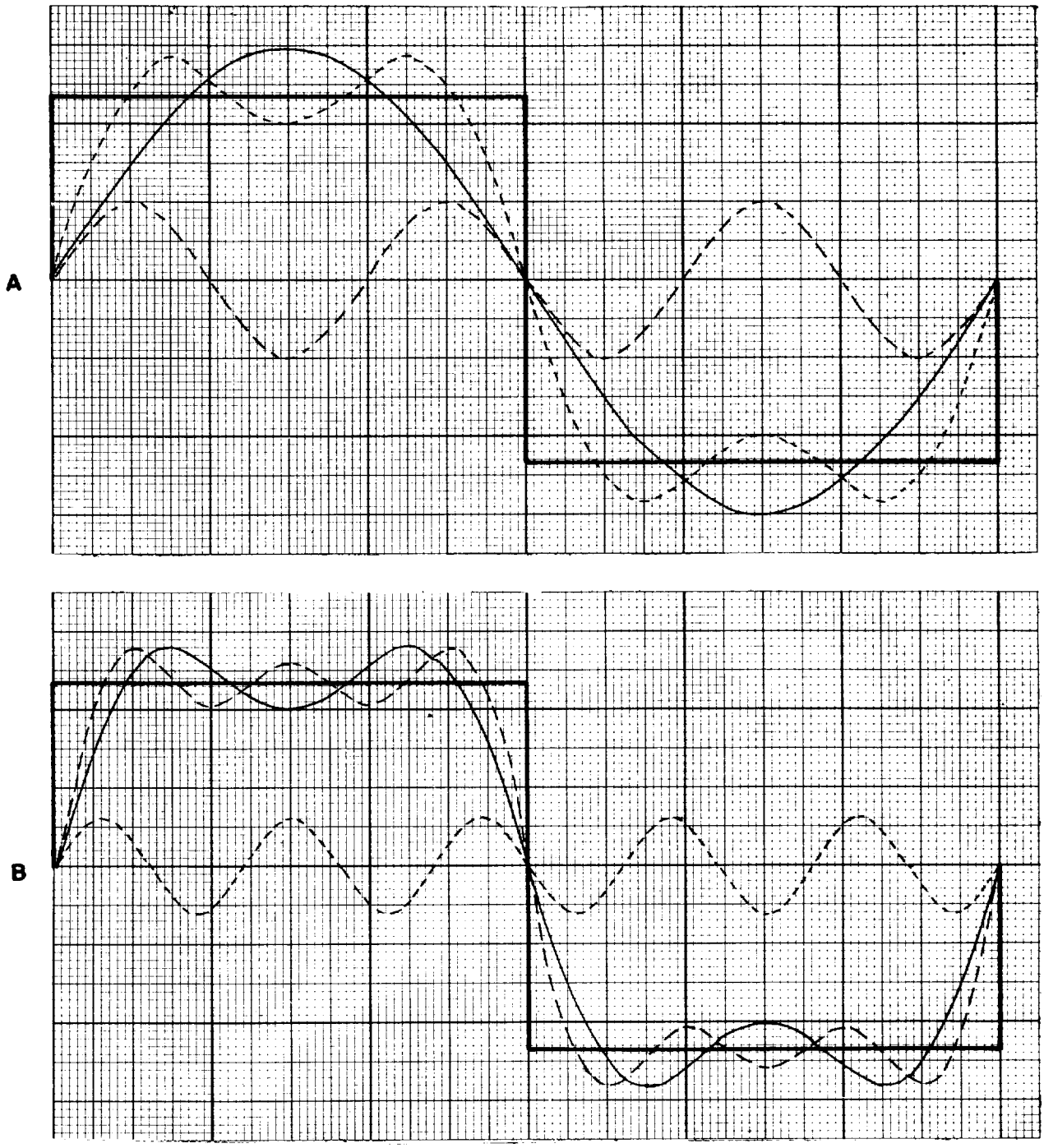


Figure 15 - Components of a Squarewave

DOCID: 3726612

at the proper points. This requires that the sinewaves have a definite relationship in frequency, amplitude and phase. If any of these relationships are altered in any way, the end result will be something other than that which is desired. For the squarewave the sinewaves are always initially in phase and are back in phase again for every 180° of the lowest frequency component.

7. Triangular Wave from Sinewaves.

If two small modifications were made on the harmonics used to construct a squarewave these same frequencies could be used to construct one type of sawtooth wave called a triangular wave. The necessary modifications are: first, to reduce the amplitude of the third harmonic to one-ninth the amplitude of the fundamental, and the fifth harmonic to one-twenty-fifth the amplitude

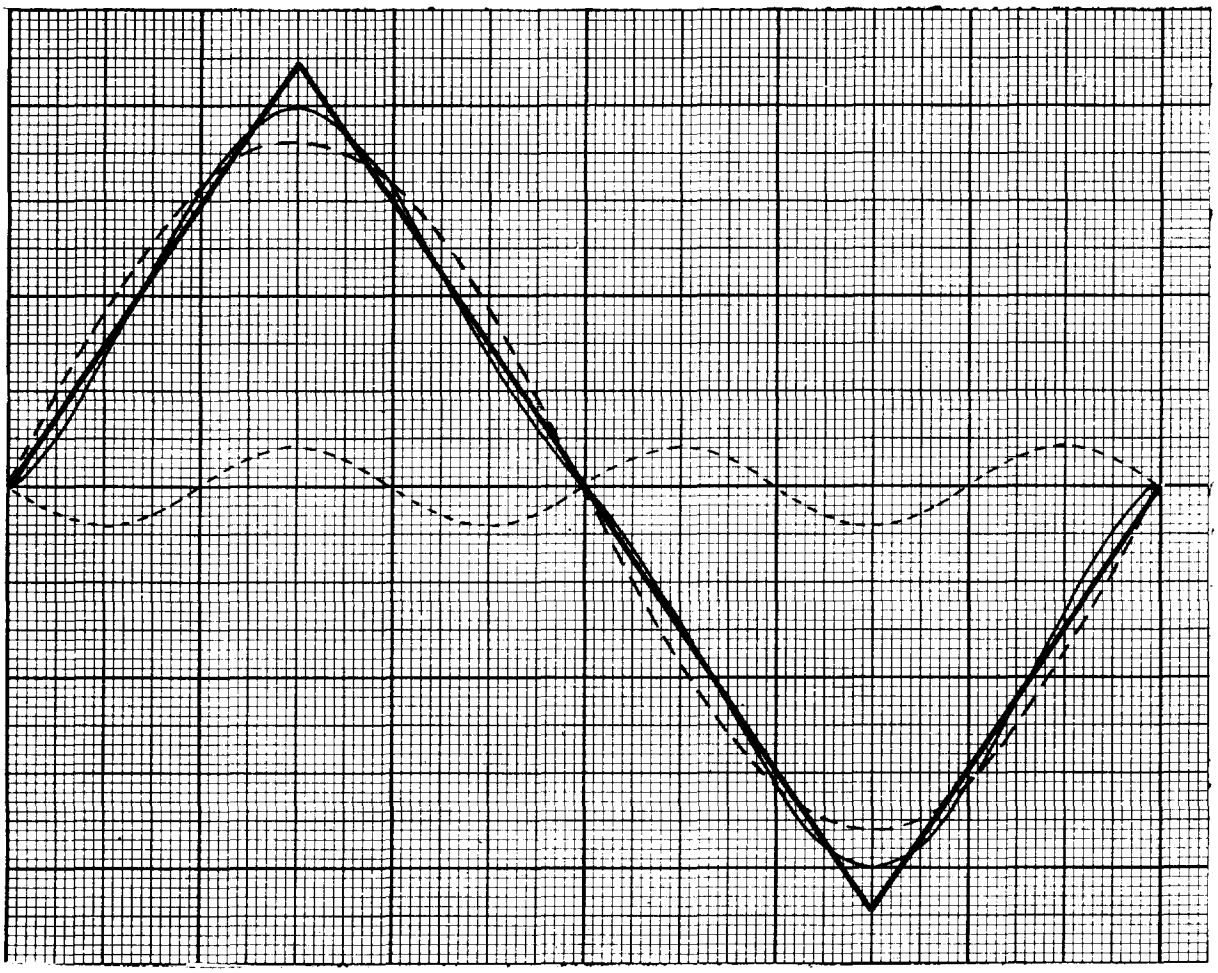


Figure 16 - Components of a Triangular Wave

of the fundamental; second, to shift the phase of the third harmonic and every second odd harmonic after it so that they are initially 180° out of phase with the fundamental wave while the fifth harmonic and every 2nd odd harmonic after it are initially in phase with the fundamental. Odd harmonics initially 180° out of phase with the fundamental are always back out of phase for each 180° of the fundamental and those initially in phase with the fundamental are back in phase for each 180° of the fundamental.

Observe that Figure 16 gives the result of just the fundamental and third harmonics. However, it is readily seen that the approach toward a triangular sawtooth is rapid. To be a perfect triangular sawtooth wave, odd harmonics extending to infinity must be used. There are, too, different shapes of sawtooth wave each of which has a different requirement as to harmonics, amplitude and phase. One frequently used sawtooth consists of all harmonics both even and odd extending to infinity.

D. Summary of Linear Mixing

When two or more sinusoidal waves of different frequency are added, a non-sinusoidal or complex wave will result. Various waveshapes may be synthesized by this method known as linear mixing or adding although for a given situation it may be much more convenient to employ some other method to acquire a desired waveshape. Linear mixing may be summarized by the following:

1. At any instant the value of the resultant complex wave is equal to the sum of the instantaneous values of the original components only.

2. No new frequencies are produced.

3. All of the original components will still be present in the same phase relationship and proportionate amplitude as they were prior to mixing.

4. None of the components undergo any distortion.

5. The original components compose the entire wave and can be extracted from the complex wave by appropriate filtering.

6. These characteristics all say, "effect is directly proportional to cause."

DOCID: 3726612

IV. FOURIER COMPONENTS

A. Classification of Waves

All waveforms may be classified under two headings, periodic and aperiodic. Under each of these headings there may be both sinusoidal and non-sinusoidal waves. Periodic simply means that the wave is a repeating function, not necessarily harmonic. Any wave, then, that repeats itself at constant regular intervals is designated as periodic. Aperiodic means that the wave is not a repeating function. Therefore, any wave, whether sinusoidal or non-sinusoidal, that occurs randomly or at irregular intervals is called an aperiodic wave. It should further be observed that although we speak of periodic waves, yet by our very definition there is no such thing since every wave has a beginning and no wave continues forever. In many practical cases, however, no complications are incurred by treating these waves as if they were truly periodic. Periodic will be used in this less conservative sense throughout the remainder of this study guide.

B. Generation of a Non-Sinusoidal Wave

The terms harmonic and sinusoidal are used interchangeably and describe a sinewave type of variation. A wave is not a sinewave if there is any deviation whatever from the sine function, therefore, a sinusoidal or simple harmonic wave when distorted in any manner is no longer a sinewave. However, as a result of distortion, a dc level and a series of harmonics are produced that are integral multiples as 1, 2, 3, 4, etc., of the original sinewave frequency. (When the distortion is caused by non-linear mixing, additional frequencies not harmonically related are also produced but a discussion of these will be deferred for more appropriate coverage under modulation.) The distorted wave is found to be various multiple harmonics depending upon the shape of the distorted wave, plus a dc level combined as if by linear mixing. The resultant wave is, by definition, non-sinusoidal. The distortion that caused this non-sinusoidal resultant makes no change in the periodicity of the wave.

C. Fourier Analysis

All non-sinusoidal periodic waves may be analyzed into a dc level plus a series of harmonic components of multiple frequency whose sum is called a Fourier series. The resolution of a non-sinusoidal periodic wave into a Fourier series is called Fourier analysis. The actual mechanics of performing such an analysis is beyond the scope of this course, but the student should recognize what is meant by the term. Examples of the Fourier series for three of the

DOCID: 3726612

waves previously discussed are included below for information only as it is not intended that the students be responsible for knowing them.

1. Squarewave:

$$a = \frac{4A}{\pi}\left(\text{Sin}\,\theta + \frac{\text{Sin}\,3\theta}{3} + \frac{\text{Sin}\,5\theta}{5} + \frac{\text{Sin}\,7\theta}{7} \ldots\right)$$

2. Triangular Sawtooth:

$$a = \frac{8A}{\pi^2}\left(\text{Sin}\,\theta - \frac{\text{Sin}\,3\theta}{9} + \frac{\text{Sin}\,5\theta}{25} - \frac{\text{Sin}\,7\theta}{49} \ldots\right)$$

3. Ideal Sawtooth:

$$a = \frac{2A}{\pi}\left(\text{Sin}\,\theta - \frac{\text{Sin}\,2\theta}{2} + \frac{\text{Sin}\,3\theta}{3} - \frac{\text{Sin}\,4\theta}{4} \ldots\right)$$

The formulas given are not complete but contain only the first few components which for many practical purposes may be considered sufficient. Actually, the Fourier series extends to infinity but the higher harmonic terms fade into insignificance as their amplitude decreases as the reciprocal of their harmonic order.

A brief explanation of the Fourier series follows using the formula of the squarewave as an illustration.

"a" is the instantaneous amplitude at any specified point along the complex wave.

"A" is the peak value of the complex wave.

"$\frac{4A}{\pi}$" is the peak value of the fundamental component.

"Sin θ " is the instantaneous angle of the fundamental sinewave at the time when "a" is measured.

"Sin 3θ " is the instantaneous angle of the third harmonic at the time when "a" is measured. (Similarly with Sin 5θ for the fifth harmonic.)

DOCID: 3726612

"3" in the denominator of the fraction $\frac{\text{Sin } 3\theta}{3}$ indicates the peak amplitude of the third harmonic as related to the fundamental component. (Similarly with the denominator 5 in $\frac{\text{Sin } 5\theta}{5}$.)

" + " indicates that each of the sinewaves begins with a positive alternation. (A "-" sign as found in the formula for a triangular wave indicates that the sinewave preceded by this sign begins with a negative alternation.)

If a dc component were present, this fact would be indicated just ahead of $\frac{4A}{\pi}$.

D. Average dc Level and dc Components

The average dc level of any periodic wave is equal to the average value of the wave over an entire cycle. If, for example, the average value of a wave over a complete cycle is zero, the average dc level is zero and the dc component is, therefore, not present. If on the other hand, the average value of a wave over a complete cycle is not equal to zero, the average dc level is not zero, and there is a dc component present. The distinction should be noted that a dc level can exist whether a dc component is present or not, however, this distinction is only apparent when the dc level is equal to zero. For this reason the terms dc level and dc component are sometimes used interchangeably although it is not always quite correct. It should further be observed that the presence or absence of a dc component has no effect on the shape of a complex wave but only its position above, below or about a zero reference level.

E. Fourier Analysis and Wave Spectra

The harmonic components that constitute a wave are called the spectrum of that wave. The importance of Fourier analysis is to discover the spectrum of a given wave. The spectra of certain standard waveforms are common knowledge and may be found under Fourier analysis in available reference books. The analysis is useful in that it shows how a non-sinusoidal wave, however complex, can be resolved into a series of sinewave components each of which must bear a definite relationship to every other component in a specific phase, multiple frequency, and proportionate amplitude. For ease, these relationships are referred to the fundamental component. The wave yielding a particular Fourier series may be reconstructed by recombining the components in the proper relationship by linear mixing. The Fourier analysis and hence the spectrum of a particular wave can be better appreciated when one realizes that this provides a knowledge of the bandwidth requirement of that wave.

DOCID: 3726612

V. BANDWIDTH AND MISCELLANEOUS COMPLEX WAVEFORMS

Theoretically, it is possible to have frequencies anywhere within a range from zero frequency or dc to infinity. This range is called a band. The band may be sub-divided into lesser bands. The extent of any band of frequencies is measured between the limits set by the lowest frequency and the highest frequency of the band. This range between the lower and upper limits is defined as the bandwidth. Frequencies occurring anywhere within a given band constitute a spectrum. From these definitions one can see that the spectrum of a pure tone or sinewave has but a single frequency and has no bandwidth. This is because for a single frequency, the upper and lower limits are one and the same and hence there is no area between these limits. Since most waves used in communications are complex, the spectra of most waves that we will treat will consist of more than one frequency. This is another way of saying that these complex waves have a bandwidth.

It has been established that for any given complex waveshape only one combination of frequency, amplitude and phase of its components will produce the desired result. Therefore, if a wave is to provide specific information, its shape must be retained and its components cannot be altered. Any alteration of the components will alter both the shape and the information content. The only way to prevent the alteration of a waveshape and loss of information is to have equipment with an adequate bandpass to handle equally all the components of that wave's spectrum.

Knowledge of the bandpass requirements for an equipment necessarily requires a knowledge of the bandwidth of the spectrum of the waves involved. This is the whole reason for our study in analysis and synthesis of complex waveforms. Helpful tools in the analysis of waves are the oscilloscope, the spectrum analyzer and the Sonograph. We have indicated before the type of presentation these equipments give in regard to sinewaves; now we may observe their presentation for complex waves. Only the oscilloscope will present a complex wave in complex form, i.e., as the resultant of its linearly mixed components. The spectrum analyzer and Sonograph separate the complex wave into its individual sinewave components or spectrum. On these latter two devices, it should be pointed out, the presence of sinewave components is represented by a straight line rather than a waveform.

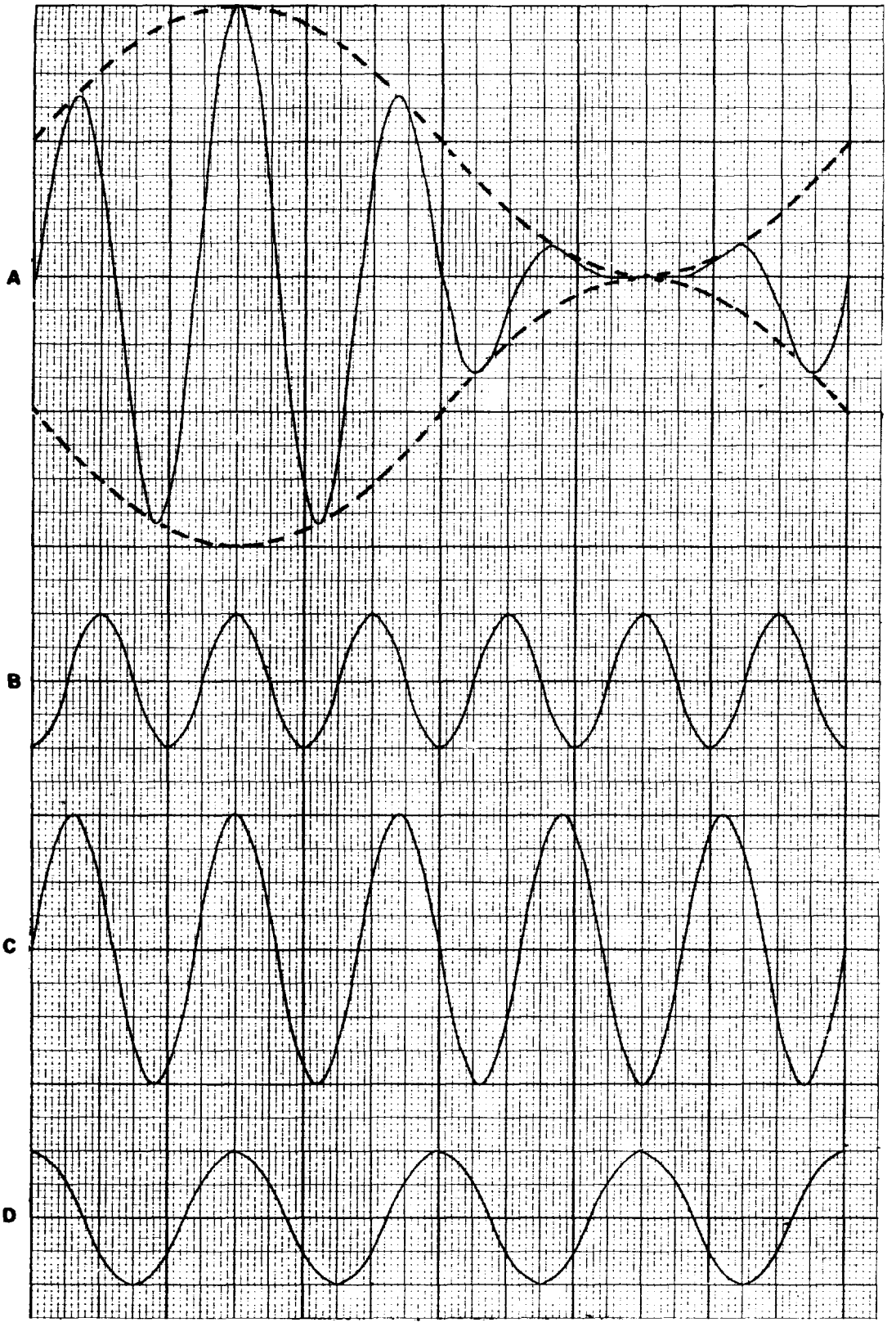


Figure 17 - A. M. Wave Components

DOCID: 3726612

A. An AM Waveform

One very important type of complex wave is that which is produced by amplitude modulation. The waveform shown in Figure 17A is the result of modulating one frequency, which is called a carrier, by a pure tone of somewhat lower frequency. The generation of this wave will be discussed later. Here, it must suffice to say that this wave which is typical of an AM wave consists of three frequencies linearly mixed. One of the three is the carrier and the others are two new frequencies, one whose frequency is equal to the carrier frequency plus the frequency of the tone and the other whose frequency is equal to the carrier frequency minus the frequency of the tone. The first of these new frequencies, the sum frequency, is called the upper side frequency and the second, the difference frequency, is called the lower side frequency.

Here, as in other complex waveforms, the components must bear a particular relationship as to frequency, amplitude and phase. Since the modulation method automatically determines this relationship, our only task is to observe what that relationship is. Wave C is the carrier. If we assume it to be a sinewave initially positive-going, the upper side frequency, B, must be a cosinewave initially positive-going and the lower side frequency, D, must be another cosinewave initially negative-going. (The cosinewave differs from the sinewave in appearance only in that the cosinewave begins with a maximum value rather than a minimum one. The cosinewave bears the same relationship to the cosine of the angle as the sinewave bears to the sine of the angle.) The complex wave shown in this figure was actually obtained by the point-by-point addition of the instantaneous values of the three components. The wave (indicated by broken lines) bordering the complex wave is the shape of the tone used to produce the modulated wave but note that this tone frequency is not a component of the resultant wave. This wave could also be developed graphically from the vectors of the individual components just as has been done with some previous waves but this will not be done here as there are complexities involved in showing the movement of three vectors in simultaneous motion.

Figure 18 shows this complex modulation wave as it would appear on each of three devices: oscilloscope, spectrum analyzer, and Sonograph. Note how the oscilloscope shows only the resultant complex wave, the spectrum analyzer device shows the individual components of the wave giving their relative frequency and relative amplitude, while the Sonograph indicates the existence of these components with respect to time.

DOCID: 3726612

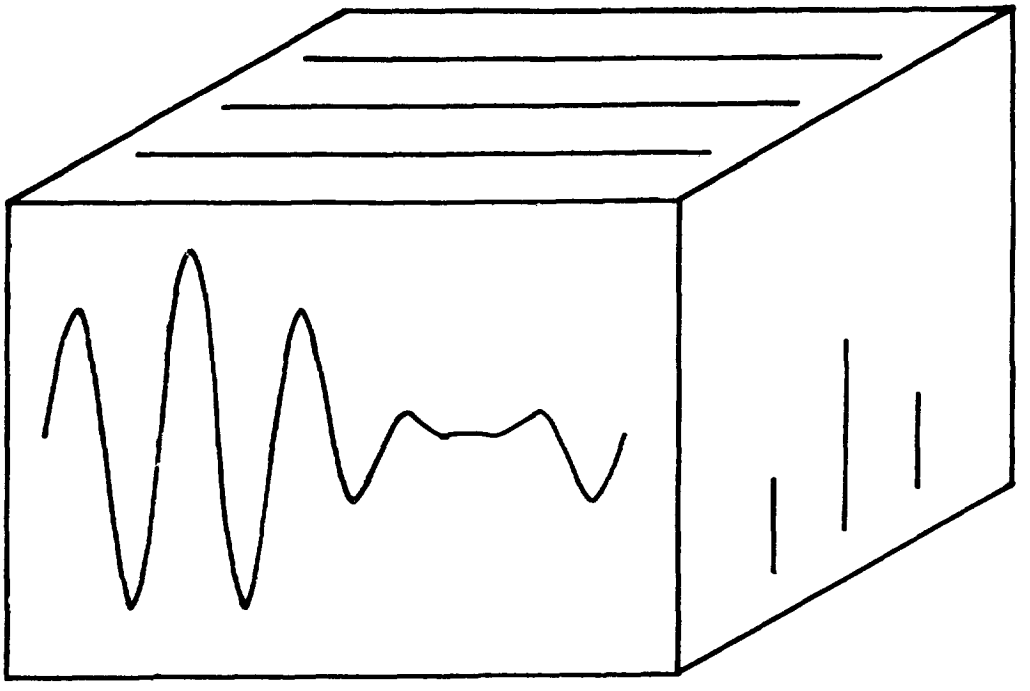


Figure 18 - Three Views of an A. M. Wave

B. A Wave with 1 kc and 5 kc Components

Suppose the complex wave illustrated in Figure 13C were to have a period of one millisecond. The frequency of the fundamental component would then be one kilocycle. Since for each cycle of the fundamental frequency there are five cycles of the other component, the latter component's frequency must be 5 kc. For convenience the complex resultant is redrawn in Figure 19A. This is the way that it would appear on an oscilloscope, but if this wave were applied to the input terminals of a spectrum analyzer the presentation of Figure 19B would appear. The same wave displayed as a sonogram would be that of Figure 19C.

C. A One Millisecond Squarewave

If the squarewave of Figure 15B were analyzed by the method already employed, the spectrum and sonogram would be those shown in Figure 20. If a dc component were present, as is often the case, its presence would be found directly along the amplitude axis of the spectrum presentation. Whether or not a dc level were present this wave's spectral lines would appear at 1 kc, 3 kc, and 5 kc.

DOCID: 3726612

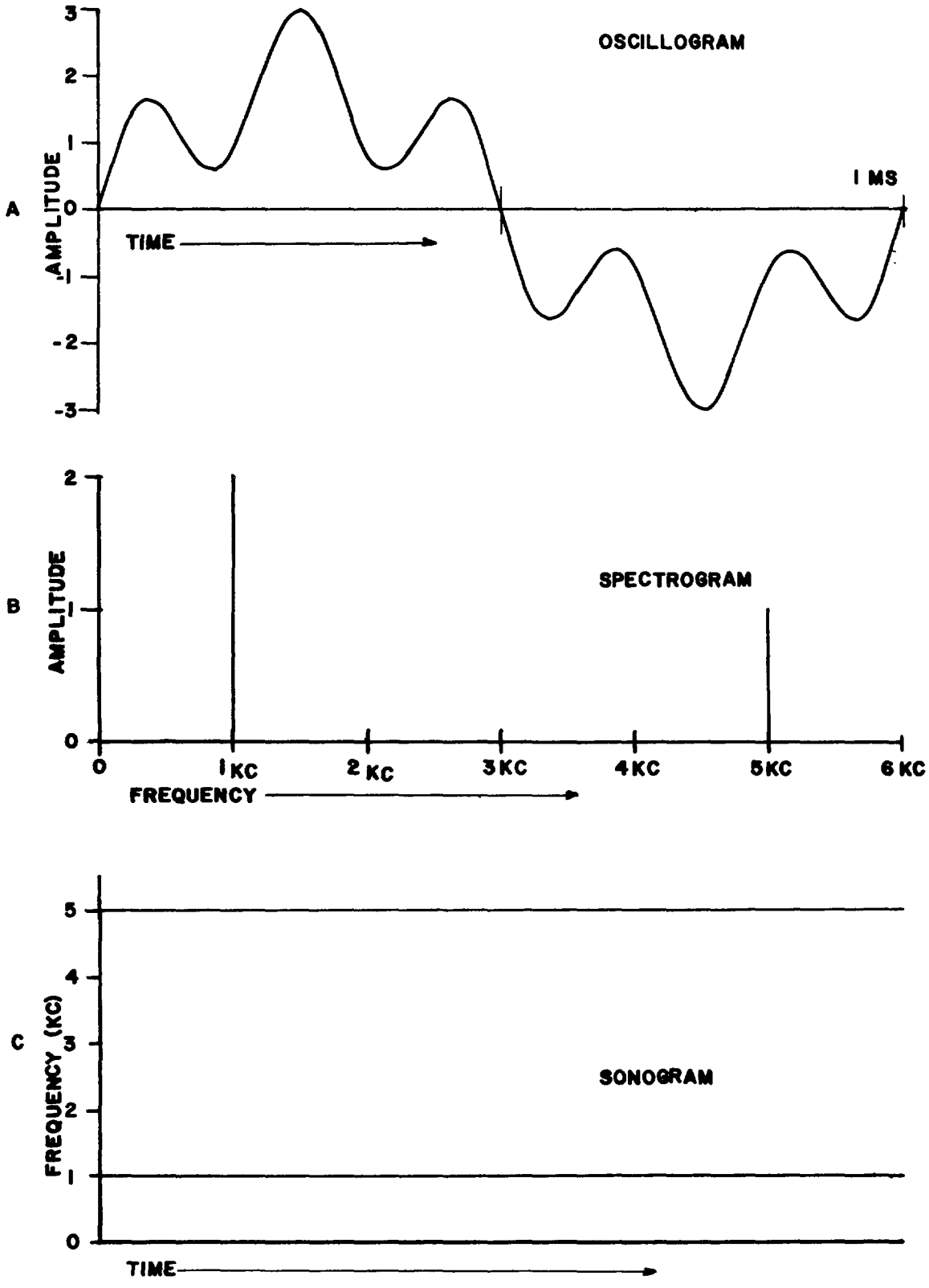


Figure 19 - Three Views of Figure 13

DOCID: 3726612

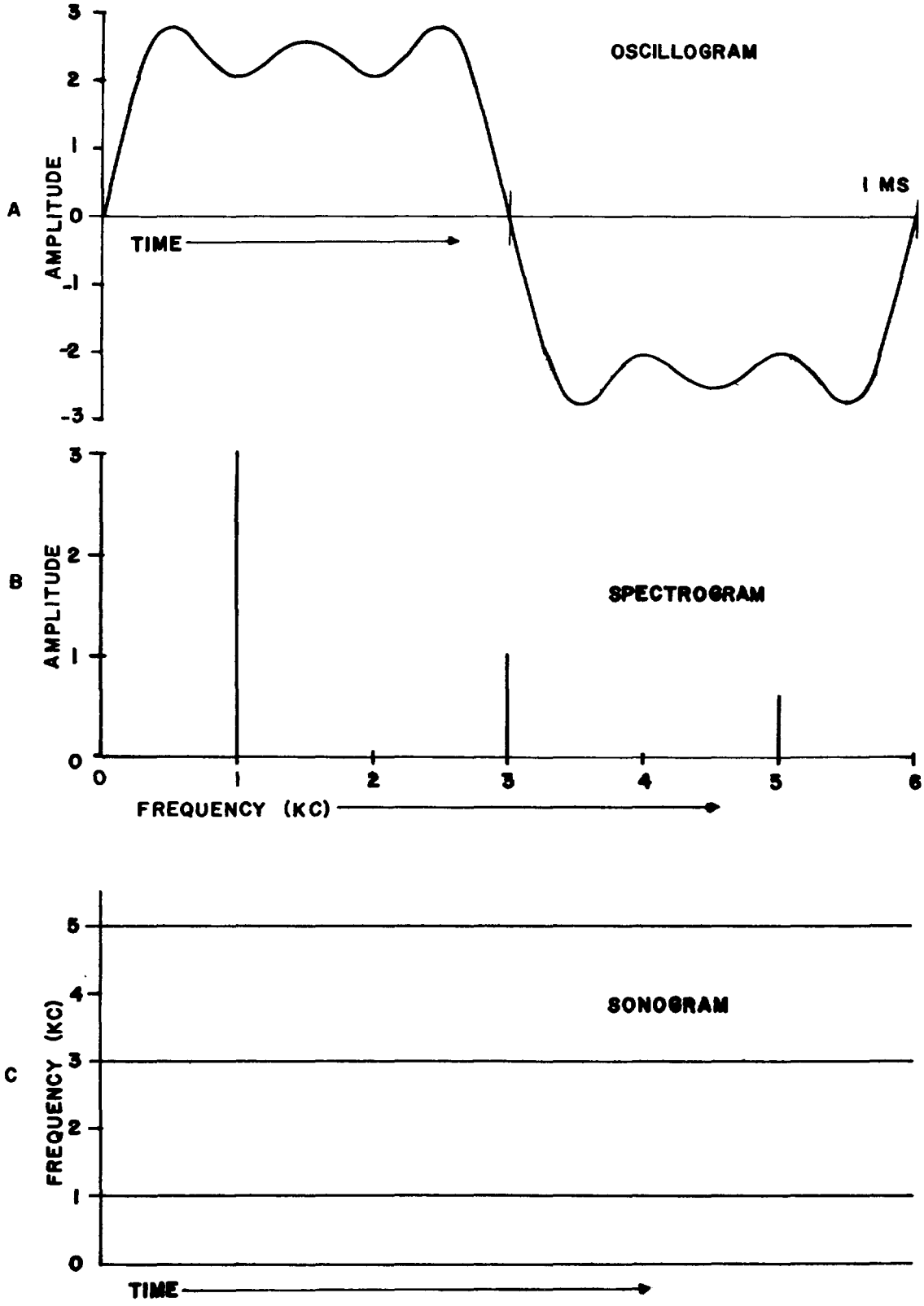


Figure 20 - Three Views of a Squarewave

DOCID: 3726612

D. A Typical Voice Wave

As typical of voice intelligence, the complex wave uttered by the human voice in pronouncing the letter "A" is shown in Figure 21. The wave as it appears on the oscilloscope is illustrated in 21A. When "A" is pronounced, it is actually done in two parts. The first part is $\overline{AA}$ and the second part is $\overline{EE}$. These two parts are in evidence in the spectrogram where a sharp division appears at about 2 kc. This is also in evidence in the sonogram. The $\overline{AA}$ sounds are lower in frequency and greater in amplitude than the $\overline{EE}$ sound. Careful pronunciation of "A" indicates the rise in frequency or pitch as the sound of the letter approaches termination. Letters whose sounds divide into two parts such as this are called diphons, a term literally signifying two voices.

The frequencies involved in normal speech vary from one individual to the next but the average speaking voice of the adult male has a fundamental frequency of about 128 cycles per second which corresponds to C below middle C on the piano. A deep voiced male may have a low of about 90 cycles per second. The average speaking voice of the adult female has a fundamental tone of about 256 cycles per second which is twice that of the average male voice and corresponds to middle C on the piano. A shrill-voiced soprano may have a fundamental pitch as high as 300 cycles per second. The range of the speaking voice is extended still further upward by two other factors. The first is the blending of overtones or harmonics up to as far as 6 kc and the second is by the sounds of certain consonants such as the hiss of the letter "S" which may go as high as 10 kc. The singing voice has an even greater range going as low as 60 cycles for a deep bass and as high as 1300 cycles for a high soprano with harmonics going up to 10 kc.

So far as bandwidth is concerned, the frequency range that is necessary in order to transmit voice intelligence depends upon the desired fidelity. Entertainment functions use the greatest bandwidth since they must not only accommodate speech and vocal music but instrumental accompaniment as well. Out of the audible range of 20 to 20,000 cycles, special music programs have a band extending from 35 to 8,000 cycles and ordinary broadcast a band of 100 to 5,000 cycles.

Pure communications functions use the least bandwidth since they are concerned with speech only. The most important frequencies of speech fall almost entirely below 3 kc. Most of the energy even in this range is below 1 kc but these tones below 1 kc contribute little to the intelligibility of the words.

DOCID: 3726612

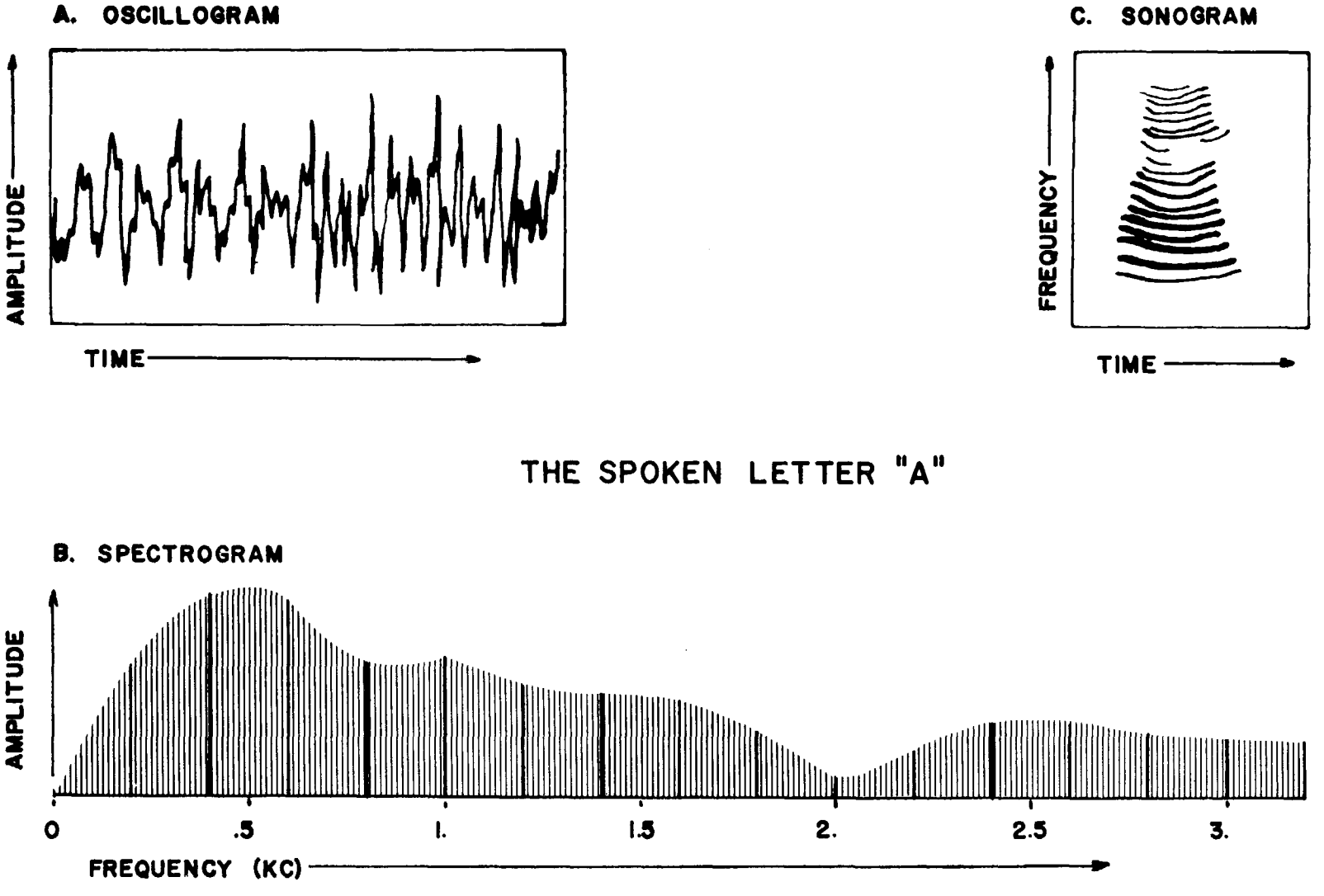


Figure 21 - Three Views of a Voice Wave

DOCID: 3726612

Conversely, the frequencies above 1 kc contain but a small percentage of the speech energy but are extremely important to the intelligibility. For example, frequencies greater than 1,000 cycles per second provide 86% of perfect articulation but only 17% of the energy. It is found that the bare minimum band for transmission of intelligible and identifiable speech is 500 to 2,000 cycles. The general average band in use is 250 to 2,750 cycles but the best current practice is to consider 100 to 3,500 cycles as the voice range. As this course develops, it will be found that specific systems will have some specific upper and lower limits and the values used here are a guide only.

CONCLUSION

In the treatment of waveform analysis continuous stress has been laid upon the sinewave, its measurement and its usage as a basic building block in the analysis and synthesis of complex waves. The accomplishment of this has introduced words and terms that will have a continuous usage in the studies that follow. The mastery of these words, terms and relationships will permit the student to proceed without appreciable difficulty.

VI. PRACTICAL EXERCISES

1. The Sonagraph presentation is a plot of

   a. Amplitude as a function of time
   b. Amplitude as a function of frequency
   c. Frequency as a function of time
   d. Frequency as a function of amplitude

2. All true sinewaves are

   a. A pure frequency
   b. Semicircles
   c. Complex waves
   d. Arbitrarily plotted

3. The spectrum analyzer presentation plots

   a. A trapezoidal wave pattern
   b. Amplitude as a function of time
   c. A sinewave pattern
   d. Amplitude as a function of frequency

4. When a sinewave is distorted

   a. The spectrum is unchanged
   b. Harmonics are generated and a complex wave is formed
   c. No analysis can be performed
   d. The fundamental frequency cannot be determined

5. Linear mixing is also called

   a. Heterodyning
   b. Adding
   c. Filtering
   d. Harmonizing

6. In linear mixing, if the phase of any component of a complex wave is changed

   a. The waveshape also remains the same
   b. The waveshape changes
   c. It can't be done
   d. New harmonics are generated

7. If two waves are in opposition they are

   a. 90° out of phase
   b. 180° out of phase
   c. in phase
   d. 210° out of phase

8. The oscilloscope is useful for measuring

   a. The energy distribution of a transient
   b. Amplitude as a function of time
   c. I. Q. of the enemy
   d. Frequency as a function of amplitude

DOCID: 3726612

9. For a specific complex wave shape (Give the best answer)

   a. Only certain harmonics can be present
   b. Amplitude of the harmonic components is important
   c. Phase of the harmonics is important
   d. Only certain harmonics can be present, and each harmonic can have only one specific amplitude and phase relationship with respect to the fundamental

10. Frequency is measured in

   a. Meters per second
   b. Meters per cycle
   c. Cycles per second
   d. Seconds per cycle

11. What is the formula for sinewaves? Explain.

12. The peak amplitude of a certain sinewave is 50 volts.

   a. What is the instantaneous amplitude at 30°?
   b. What is the instantaneous amplitude half a period later?
   c. Assuming the wave starts with a positive alternation, what would be the polarity of the instantaneous amplitude found in part b?

13. Differentiate between frequency, period and phase.

14. What determines the name of a particular harmonic?

15. Define a "periodic" wave.

TRUE OR FALSE

16. A distorted wave will produce sub-multiples as well as harmonics.

17. The average dc level of any periodic wave is equal to the average value of the wave over an entire cycle.

18. The two general classes of waves are sinusoidal and non-sinusoidal.

19. A dc level cannot exist if a dc component is not present.

20. The shape of a complex wave is directly due to the presence or absence of a dc component.

DOCID: 3726612

CHAPTER 2

PULSE ANALYSIS

An earlier study in this text introduced an elementary approach to waveform analysis. The ideas that were put forth will be extended here to apply to a particular kind of wave which is often aperiodic or transient and which is classified under the general heading of pulse. Pulses find wide application in telemetry, teletype, television, telegraph, computers and radar. In these applications and others, pulses provide a means of conveying some type of information. Described by the American Standards Association as, "a brief excursion of a quantity from normal," pulses have various shapes and forms, each particular shape having its own characteristic spectrum and a minimum bandwidth requirement which must be met if the ideal waveshape is to be approximated by the actual pulse shape. Methods for generating pulses range from simple to complex depending upon the needs of a specific application, upon the general shape that is desired, and upon the degree of perfection to which the desired shape must be matched. Irrespective of application or how generated, realization and retention of a waveshape is dependent upon the ability of the equipment to pass the pulse spectrum with equal treatment of all components. Hence a study of pulse applications and pulse shapes is incomplete without a study of pulse spectra since equipment bandwidth must always be wide enough to accommodate the bandwidth of the signal employed. The present lesson is intended to broaden the student's appreciation for signal bandwidth and equipment bandwidth.

I. PULSES

Morse telegraph, digital computers, radar, television, facsimile, and teletype all owe their success, and in some instances even their existence, to a peculiar type of electrical wave called a pulse which may be thought of as a short burst of energy before and after which there is a rest. In these systems pulses are used either to carry intelligence, or to control equipment, or both. To better understand pulses it is first desirable to find a definition, then to note some of the types that are widely used and finally to observe the bandwidth requirements for transmitting them with useful fidelity.

A. Definition

Imagine an electrical circuit equipped with a key or switch and supplied by a dc source such as a battery. The switch has been open for a long time. The switch is then closed and quickly reopened. At the instant the switch is closed, the voltage applied across the circuit is suddenly stepped from "zero" voltage to the full battery voltage. At the instant the switch is reopened, the voltage applied across the circuit is stepped from the full battery voltage back to zero. A sudden, short change in the level of voltage (or current) such as this constitutes an electrical pulse.

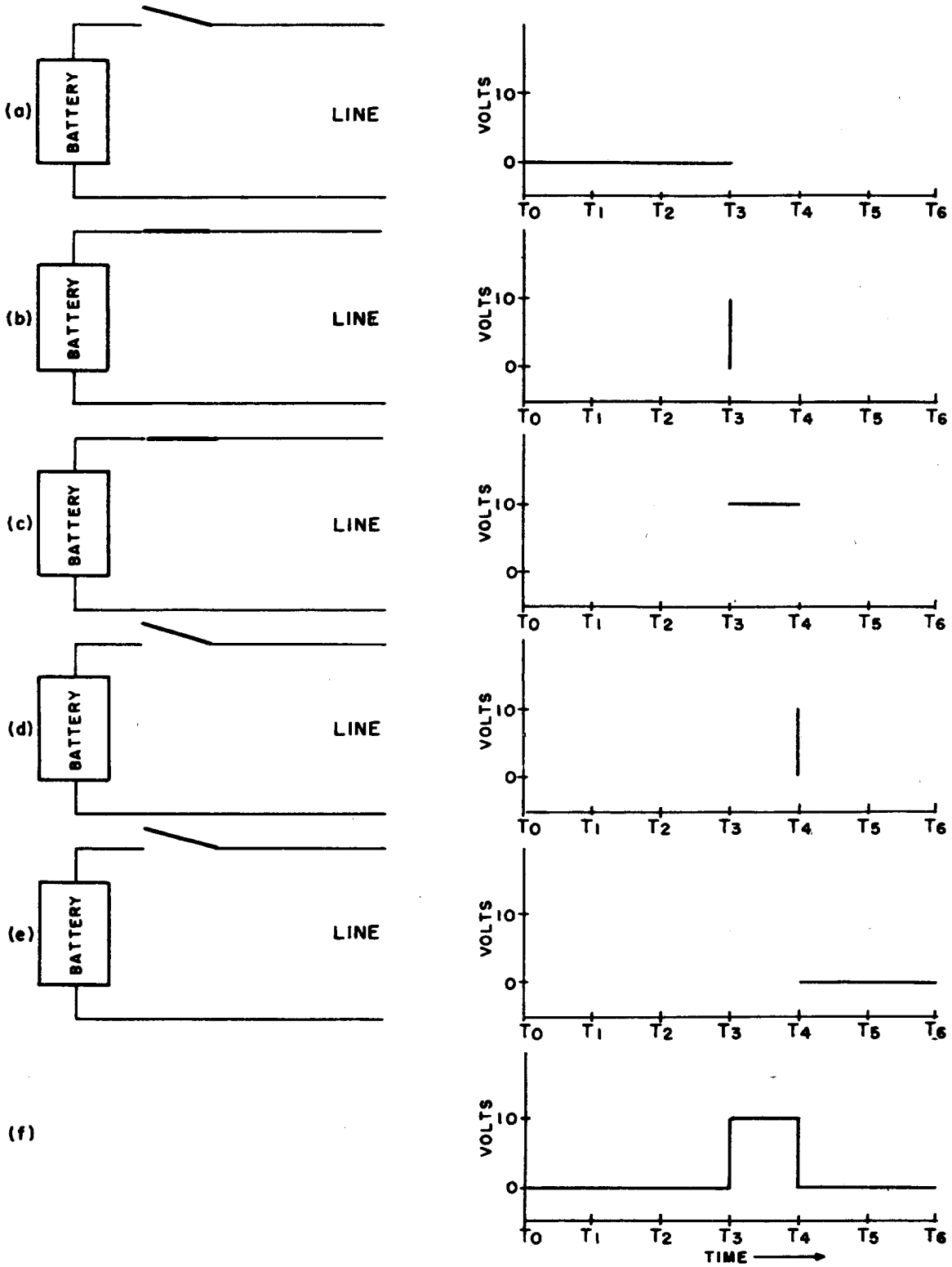


FIGURE 22. GENERATION OF A PULSE

DOCID: 3726612

The sequence just described is indicated in Figure 22. In 22A the switch is shown open for a "long time" indicated by the interval from time T_0 to time T_3. The abrupt rise or step to the new level when the switch is closed is shown occurring at time T_3 in Figure 22B. The new level is maintained from times T_3 to T_4, (Figure 22C), and then at time T_4 a second step returns the voltage across the circuit to its original value (Figure 22D). Figure 22E then displays the original conditions continuing just as they had before these conditions were interrupted. A composite drawing of the occurrences indicated in Figures 22A, 22B, 22C, 22D, and 22E is given in Figure 22F where it becomes apparent that the pulse is simply an interruption in an otherwise constant condition.

All of this preparation finally leads us to the actual definition of a pulse as given in the standards of the Institute of Radio Engineers, 51 IRE 20.S1, which states that a pulse is "... a variation of a quantity whose value is normally constant; this variation is characterized by a rise and a decay, and has a finite duration." It will be seen that the description previously given conforms to this definition.

B. Pulse Types

Electrical pulses have various shapes among which are rectangular, square, sawtooth, double, exponential, stepped, cosine squared, and bandwidth limited. In fact, about the only restrictions on the variety of pulse shapes are those of need for a given pulse shape, and those imposed by the ability of circuits to produce them. Some of the standard pulse shapes in wide use in electronics and communications are illustrated in Figure 23.

Methods employed in generating various shapes vary widely. Typical methods for rectangular and square waves are either to switch alternately between two dc voltage levels or else to clip the positive and negative peaks from a sinusoidal wave. In turn exponential pulses may be derived from a square or rectangular pulse by passing the wave through a simple low-pass filter. Some pulse shapes are obtained by linear mixing of other waves that have the necessary relationship to each other in regard to time and amplitude. Still others are obtained by direct generation as with the linear time-base or sawtooth waveform.

1. Rectangular and Square Waves

Throughout this text reference is made to rectangular and/or square waves. In the strict geometrical sense neither of these exists because limitations imposed by circuitry prevent the realization of geometric perfection, but if these limitations which take the form of finite rise and decay time (to be discussed) are disregarded, then any wave which alternately assumes one of two fixed values belongs to the general classification of rectangular waves.

DOCID: 3726612

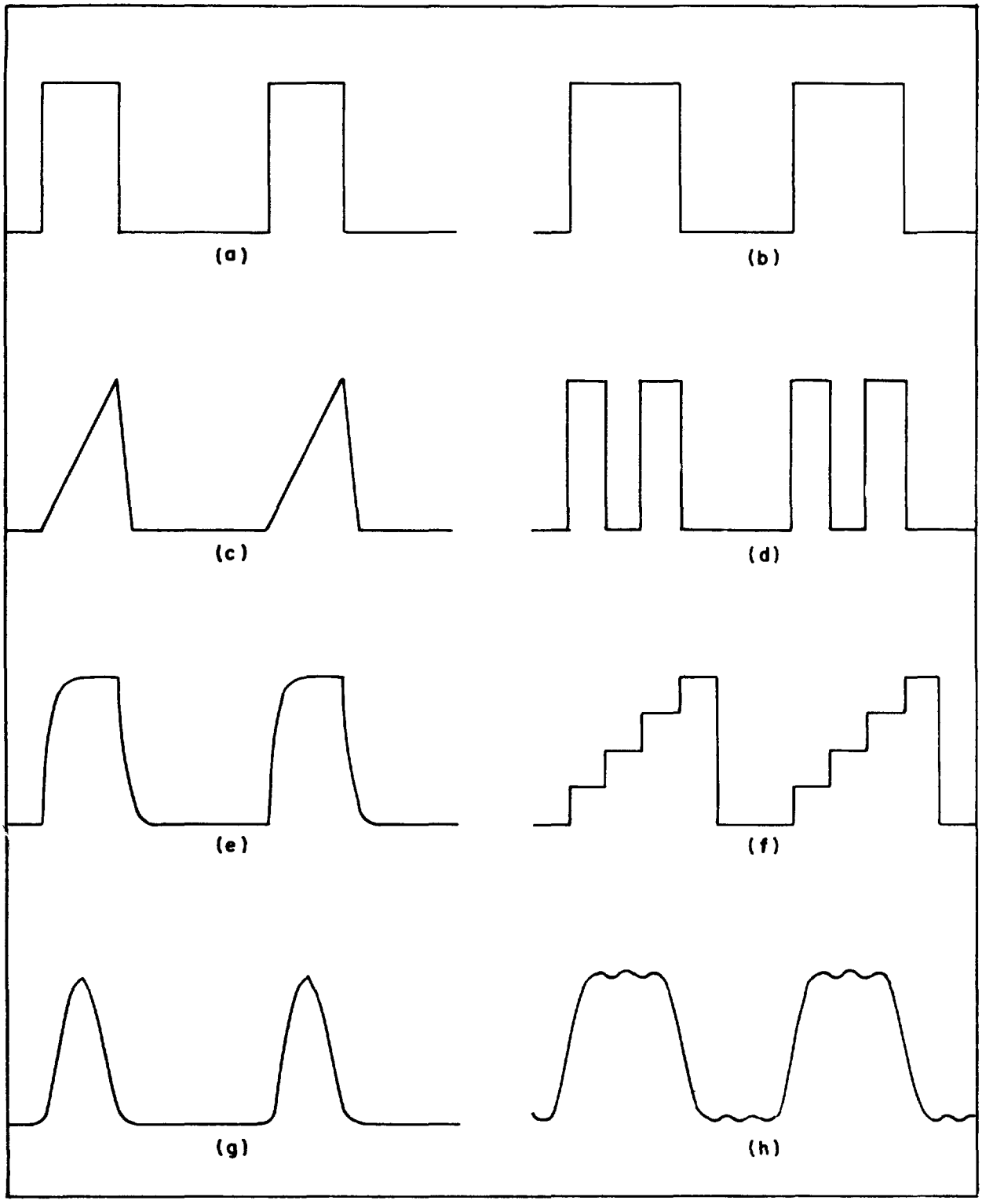


FIGURE 23. SOME TYPES OF PULSES.

DOCID: 3726612

In geometry all squares are rectangles but not all rectangles are squares. Similarly, with regard to pulse-type waveforms, an extension of this reasoning allows it to be said that all squarewaves are rectangular waves but not all rectangular waves are squarewaves. Of the infinite number of time ratios resulting in rectangular waves, the squarewave is unique, characterized by its equal mark and space elements. Figures 23A and 23B are both rectangular waves but, specifically, Figure 23B is a squarewave.

2. Clipped Sawtooth Pulse

Another pulse shape with which the student should already be familiar is that of the sawtooth (Figure 23C). It is the type of pulse that is used to deflect the electron beam across the oscilloscope screen so as to produce a trace that sweeps the screen linearly with respect to time. The pulse is seen to gradually rise to a peak and then rapidly decay to its base value. Both the rise and decay are shown to be linear with respect to time, although practical application does not necessarily require that the decay be linear. When this type of pulse is used for an oscilloscope sweep voltage, the beam is visibly moved from left to right during the gradual rise and then retraces (usually invisibly) from right to left during the rapid decay.

3. Double Pulse

Time division multiplexing takes samples of information from some specified number of sources and then transmits these samples as a sequence of pulses, each pulse representing a sample from a different source. After all sources have been sampled once the sequence is repeated. Each such complete sequence of samples is called a frame. To recover the data it is necessary to distinguish when one frame ends and the next begins. This is achieved by a pulse code acting as a "frame marker" or synchronizing pulse. The pulse code simply imparts to the frame marker a characteristic that makes it unique in the pulse train. The characteristic may be one of a special amplitude, duration, spacing or number. One possible marker pulse that can be used to distinguish the beginning of each frame is the double pulse shown in Figure 23D.

4. Exponential Pulse

An example of the more common forms of distortion of a square or rectangular wave is Figure 23E. Rectangular pulses that have been subjected to the loss and phase shift of the higher harmonics tend toward

DOCID: 3726612

the exponential pulse. This is characterized by the exponential rise of the leading edges and the exponential decay of the trailing edge. Rectangular waves passed through a simple R-C filter will take on this characteristic. The student is referred to Methods of Modulation, pages 77 and 78,for examples of the deterioration of rectangular pulses into exponential pulses.

5. Stepped Pulses

Some electronic applications require that recurring rectangular pulses be counted. One means of doing this is the use of a step-counter. This arrangement provides a voltage step, all subsequent steps being greater in amplitude than the last step. The outcome is a wave having the appearance of a staircase which rises to some predetermined amplitude after which it recycles to generate a new staircase or stepped pulse. The waveform of such a system is illustrated in Figure 23F.

6. Cosine Squared Pulse

The cosine squared pulse illustrated in Figure 23G has the significant characteristic of requiring but a very narrow bandwidth for the passage of its important harmonics. This is because most of its harmonic energy is concentrated into a very narrow band, the balance of the harmonics having negligibly small amplitude. Since this type pulse has such minimal requirements, it has been brought into usage in some modern communications systems, an example of which is the JANET meteor-burst system. An increase in the popularity of the cosine squared pulse may well be expected.

7. Bandwidth Limited Pulse

As the exponential pulse may be thought of as a rectangular pulse that has undergone loss in a simple R-C filter, so the bandwidth limited pulse may be considered to be the effect of a sharp cut-off filter on a rectangular pulse. In the case of the bandwidth limited pulse, the loss of harmonics is apparent but phase shift is not in evidence. The student is referred to Methods of Modulation, pages 77 and 78, for examples of the deterioration of rectangular pulses into bandwidth limited pulses.

II. PULSE NOMENCLATURE AND MEASUREMENT

At this time it is appropriate to consider the terms used in describing pulses, parts of a pulse, and pulse measurements. This discussion will be confined to

DOCID: 3726612

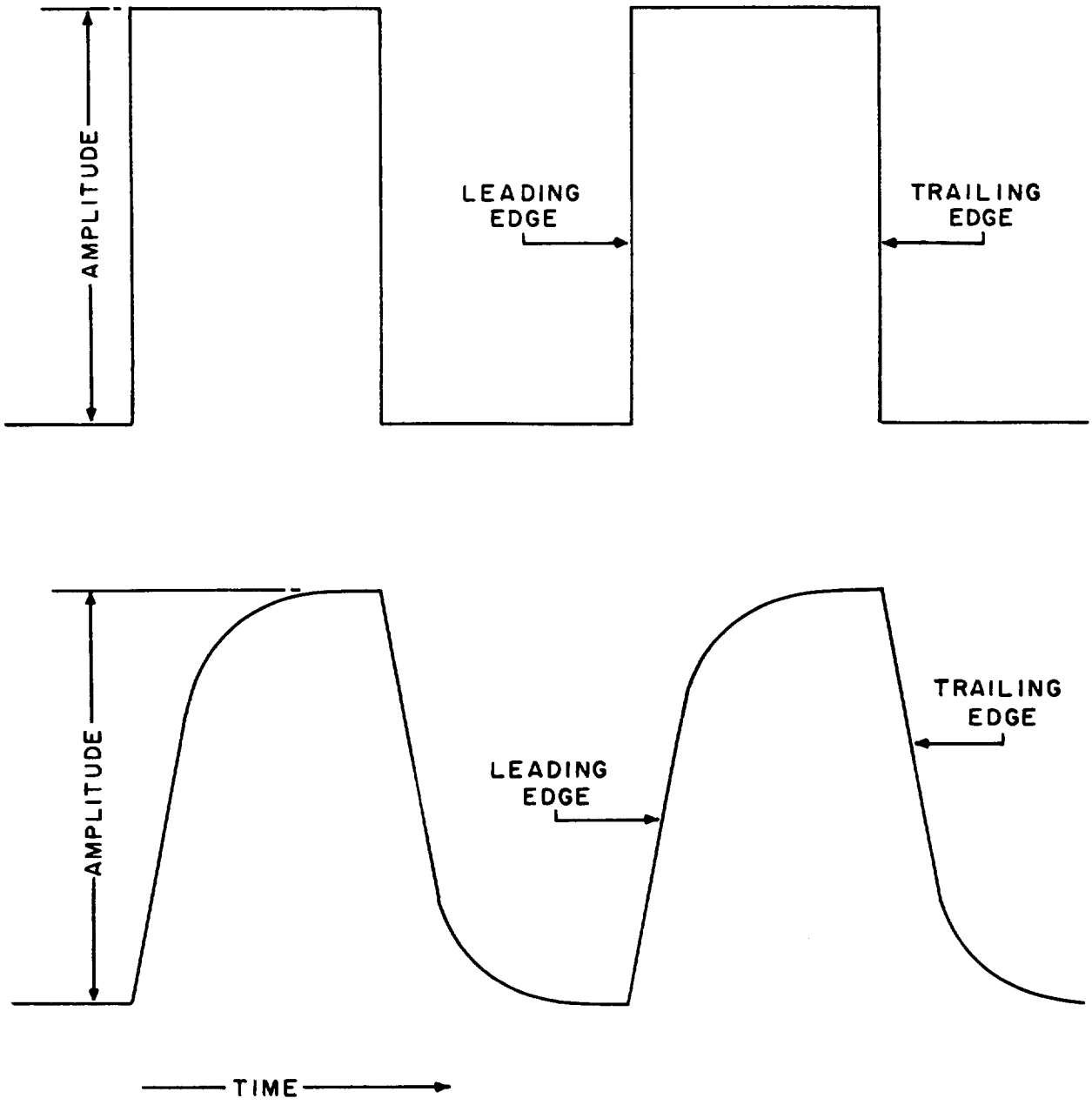


FIGURE 24. PARTS OF RECTANGULAR PULSES.

usages that apply to rectangular and square waves. The restriction is made because this type of wave is most frequently encountered in communications.

For purposes of illustrating pulse measurements the squarewave is shown in two extreme conditions in Figure 24. Figure 24A presents the ideally perfect square pulse. For reasons to be discussed this pulse is never quite realized in practice. Figure 24B approaches an exponential pulse which may be considered to be what is left of a squarewave after severe distortion resulting from phase shift and loss of high frequency components. The well-formed squarewave encountered in actual practice, as has been said, is not perfectly square; however, it is not distorted to the extent shown in Figure 24B, but is a compromise lying somewhere between the two extremes.

A. Pulse Amplitude

Every wave has some amplitude. For an electrical wave this may describe its voltage, current or power. In Figure 24 the amplitude of the pulse is illustrated as the height as measured from the base of the pulse to its top. This is the peak amplitude of the pulse. By definition the peak amplitude is the absolute amplitude of the pulse, measured from peak to peak. When speaking of pulse amplitude the peak amplitude is always intended unless some other amplitude such as average or RMS amplitude is specified.

In this writing A will be used to abbreviate peak amplitude and A_o will abbreviate the average amplitude.

B. Leading Edge

The major portion of the rise of a pulse, as indicated in Figure 24, is termed the leading edge.

C. Trailing Edge

The major portion of the decay of a pulse, as indicated in Figure 24, is termed the trailing edge.

D. Pulse Rise Time

Associated with the leading edge of rectangular pulses is the rise time. The "perfect" rectangular pulse (Figure 25A) has no rise time but all actual pulses, being something less than perfect, require some finite time to build up from their minimum to their maximum value (Figure 25B). Close note should be made that not all of the time during which the pulse rises is termed rise time. Rise time is

DOCID: 3726612

defined as the time required for the leading edge of the pulse to rise from one-tenth of its maximum value to nine-tenths of its maximum value.

E. Pulse Decay Time

Decay Time is associated with the trailing edge of the pulse as shown in Figure 25. Again, as with rise time, a "perfect" square or rectangular pulse has no decay time, but real pulses do require some time during which they fall from their maximum to their minimum value (Figure 25B). Similar to the previous definition, decay time is not defined to include the total time during which the amplitude of a pulse falls from one value to the other. Decay time is defined to include the time for the trailing edge of a pulse to fall from nine-tenths of its maximum value to one-tenth of its maximum value.

F. Pulse Duration

Pulse width and pulse duration are two terms that may be used interchangeably inasmuch as they both refer to the same thing, namely, the duration of a pulse in terms of time. Both of these terms have held preference at one time or another. This writing, however, will use the term pulse duration exclusively while using W for its mathematical symbol to be in conformity with the usage in Methods of Modulation.

1. Where Pulse Duration is Measured

In looking at the perfect square pulses in Figure 25A one can see that it would make little difference whether the pulse duration were measured near the top, near the bottom, or somewhere in between. With an imperfect pulse as would be encountered in a practical situation, however, it would make a great deal of difference where the pulse duration is measured. Observe that if it were measured at or near the bottom the measurement would be large, and if measured at or near the top the measurement would be small. In fact the pulse has some width at any point along its amplitude and any of these could properly be called pulse duration.

2. How Pulse Duration is Specified

For the reason just shown, it is necessary to specify where the pulse duration is measured, otherwise the term is meaningless. Therefore, pulse duration is defined as "the time interval between the first and last instants at which the amplitude reaches a stated fraction of the maximum pulse amplitude." In the light of this definition it can

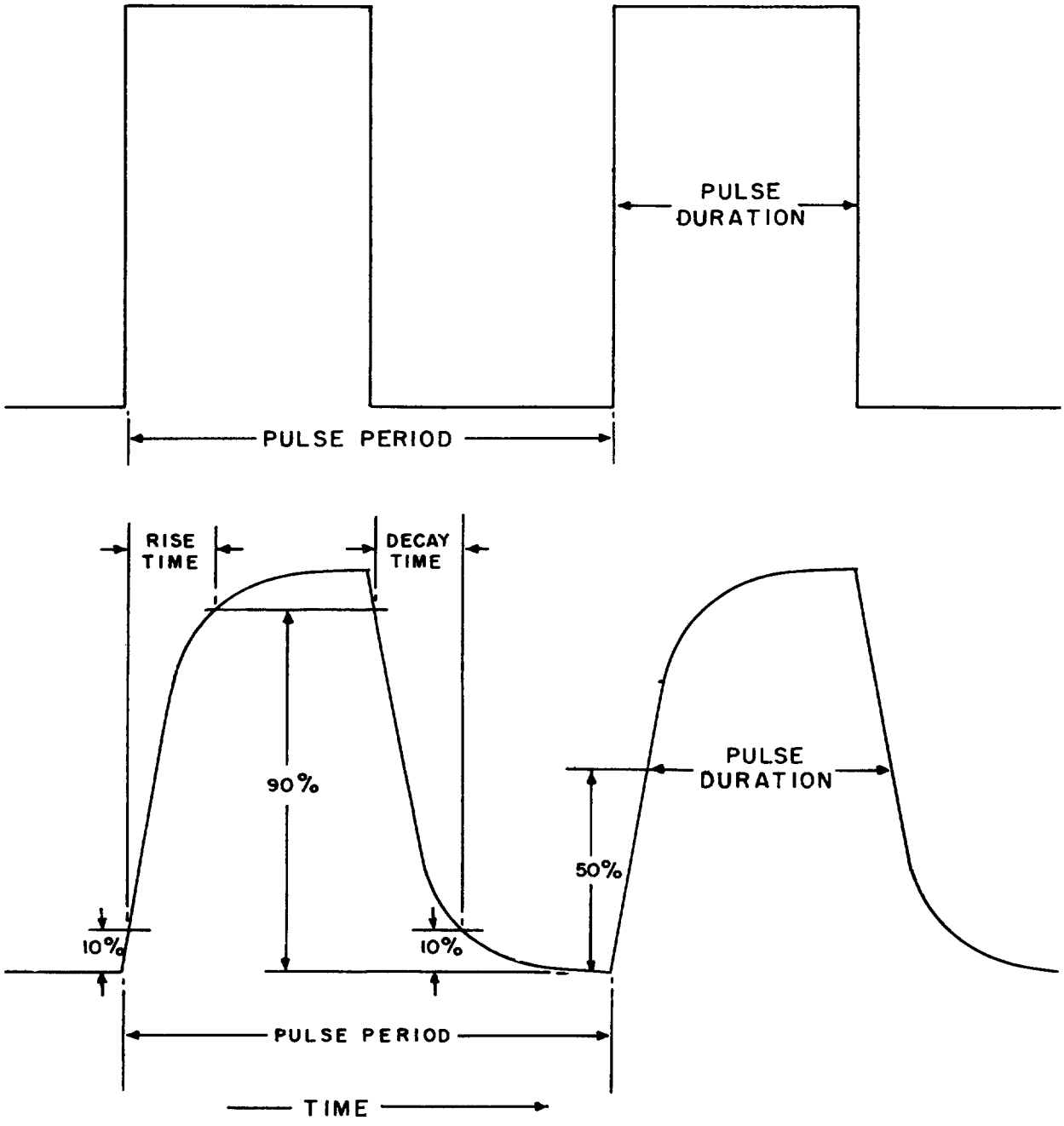


FIGURE 25. PULSE MEASUREMENTS

DOCID: 3726612

be said that the pulse duration at 75% of peak amplitude is such and such a value, or the pulse duration at 30% of peak amplitude is some other value.

3. Where Pulse Duration is Normally Measured

Upon making a comparison of the ideal and distorted pulses in Figure 25, the only level on the distorted pulse where the duration is identical to that of the undistorted pulse is at 50% of the maximum amplitude. For this reason 50% is generally accepted in current practice as being the ideal "stated fraction of the maximum pulse amplitude" for measuring the duration of a square or a rectangular pulse. In this writing pulse duration will always mean the time width of the pulse at 50% of maximum amplitude unless some other fraction is stated.

G. Pulse Period

In the chapter on Waveform Analysis, wave period was discussed briefly in regard to the sinewave. It was described as the time required to complete a cycle or the time for the sinewave to complete 360°. This would allow the period to be measured from crest to crest, trough to trough or between any other pair of identical points on adjacent alternations. It could be summarized that regardless of the points chosen for the measurement, the period is the time that elapses before the wave repeats itself. Similarly, the period of the pulse train illustrated in Figure 25 may be measured between identical points on adjacent leading edges as indicated, or between identical points on adjacent trailing edges, or between any other pair of identical points on adjacent waves.

It should be noted that some authors refer to pulse period as pulse repetition time (PRT), pulse recurrence time (PRT) and pulse repetition period (PRP). In this writing, the shorter expression, pulse period, will be used and for mathematical usage it will be abbreviated, P, just as has been done for all other wave periods.

H. Pulse Repetition Frequency

The word frequency by itself cannot be used properly for any wave configuration except sinusoidal or cosinusoidal since all complex waves including pulses consist of a number of frequencies rather than a single frequency. It is permissible, however, to describe the rate at which a periodic pulse repeats itself as the pulse repetition frequency (abbreviated PRF). PRF is given in pulses per second in much the same way that the frequency of a sinewave is given in cycles per second.

DOCID: 3726612

The student should also be aware that pulse repetition frequency is sometimes called pulse repetition rate (PRR), but the latter terminology will not be used in the present writing.

I. The Relationship of Pulse Period and PRF

The pulse repetition frequency and the pulse period are not independent of each other. Just as the frequency of a sinewave can be found by taking the reciprocal of the sinewave period, and conversely the sinewave period can be calculated from the frequency, so the reciprocal of the PRF yields the pulse period and the reciprocal of the pulse period yields the PRF. This is true of all periodic pulses regardless of their shape. These statements can be summarized by the formulas compared below:

SINEWAVES	PERIODIC PULSES
$P = \frac{1}{f}$	$P = \frac{1}{PRF}$
$f = \frac{1}{P}$	$PRF = \frac{1}{P}$

One other point that should not be overlooked is that the fundamental component of the complex wave and the PRF occur at the same rate (i.e., there are as many cycles per second in the fundamental component as there are pulses per second in the periodic pulse train) therefore the reciprocal of the pulse period supplies the identities of both. This may be stated briefly, --

$$\frac{1}{P} = f_{\text{fundamental}} = PRF$$

J. Pulse Duty Cycle

A very useful relationship of pulse measurements is the comparison of the pulse duration to the pulse period. For theoretically perfect rectangular pulses which have neither rise time nor decay time (Figures 24A and 25A) this relationship can be expressed as the ratio of the ON-TIME to the sum of ON-TIME plus OFF-TIME (i.e., the ratio of ON-TIME to TOTAL TIME). Because of the delay in the rise and the delay in the decay of pulses that occur in actual practice the ratio $\frac{\text{ON-TIME}}{\text{TOTAL TIME}}$ does not lend itself to accuracy. Improved accuracy can be obtained by using the ratio, $\frac{\text{Pulse duration}}{\text{Pulse Period}}$.

DOCID: 3726612

The previous ratio, abbreviated $\frac{W}{P}$, may be expressed in any of three ways: a common fraction, a decimal fraction, or as a percentage. The relationship is named in three ways also, to correspond to each of the ways in which it may be expressed. These are:

pulse duty ratio = fraction

pulse duty factor = decimal

pulse duty cycle = percentage

The terminology is generally used loosely, however, and the term duty cycle is frequently applied to any of the three forms. Regardless of the form in which it appears or the particular name used, the pulse duty cycle simply indicates what portion or percentage of the pulse period is occupied by the pulse width.

It should be recalled here that the pulse period is the reciprocal of the pulse repetition frequency. It should also be recalled that dividing a number by a quantity is exactly the same as multiplying the number by the reciprocal of that quantity. For this reason the pulse duty cycle can be expressed by either the ratio of the pulse duration to the pulse period ($\frac{W}{P}$) or by the product of the pulse duration and the pulse repetition frequency (W x PRF). Summarizing,

$$\text{duty cycle} = \frac{W}{P} = W \times PRF$$

1. Squarewave Duty Cycle

The squarewave is distinguished from all other rectangular waves by its duty cycle. Regardless of what the squarewave period may be, the pulse duration will be equal to one-half of this value. This automatically gives any squarewave a 50% duty cycle.

For example, assume a rectangular wave whose period is 500 milliseconds and whose pulse width is 250 milliseconds. The duty cycle is $\frac{W}{P} = \frac{250 \text{ ms}}{500 \text{ ms}} = .5$. In terms of a percentage this is 50%. Therefore this particular rectangular wave is identifiable as a squarewave by virtue of its duty cycle.

2. Peak and Average Amplitudes Related by Duty Cycle

The average value (amplitude), A_O, of a rectangular pulse is related to its peak value, A, by the duty cycle. If the peak value is known the

DOCID: 3726612

average amplitude can be found very easily by multiplying the peak amplitude by the duty factor. For example, take a 100 volt square pulse. Its duty factor must be .5 because it is square. Therefore, the average value, A_o, is .5 x 100 volts which is 50 volts. This is both the average value of the pulse and the value of its dc component.

Take a second example where there is a rectangular pulse whose period P = 100 usec, and whose pulse duration W = 75 usec. Let the amplitude again be 100 volts. The average value of the pulse and its dc component is given by --

$$A_o = A \frac{W}{P}$$

$$A_o = 100 \times \frac{75 \text{ usec}}{100 \text{ usec}} = 75 \text{ volts.}$$

III. PULSE SPECTRA AND BANDWIDTH

It will be recalled that in Waveform Analysis a number of complex waveforms were synthesized by the addition (linear mixing) of sinewaves of different frequencies. Later it was shown that periodic complex waves when analyzed by Fourier analysis can be resolved into a series of sinewaves, all of which are harmonically related. Pulses are no exception; if they are periodic they may be resolved into an average dc level and an infinite number of harmonically related sinewave frequencies, each pulse shape having a unique characteristic spectrum.

The previous remark implies that for a given wave shape the components are precisely regulated as regards frequency, amplitude and phase. On the other hand, if sinewaves of specified frequency, amplitude and phase are added, only one specific wave shape can possibly evolve.

In regard to signal bandwidth, the finite bandpass of any equipment is going to place a severe limitation on any wave that consists of an infinite number of frequencies. The question arises as to just how many frequencies out of an infinite quantity must be passed by the equipment without doing serious damage to the complex waveform. It will be the objective of the balance of this lesson to determine the necessary bandwidth requirements for pulses that are square or rectangular in form.

A. The (Sine X)/X Curve

By means of the peak amplitude, A, the pulse duration, W, and the pulse period, P, which, of course, consists of the pulse duration plus the rest time until

DOCID: 3726612

the next pulse, other important values relating to the pulse frequency spectrum may be determined. These other values are the average dc level, (A_o), the PRF, and the fundamental frequency component. Using a very interesting mathematical curve called a "sine of X over X" curve it is possible to extend the use of these values to determine the peak amplitude of the fundamental component and of each of the harmonics in the wave. The application of the "sine of X over X" curve [hereafter expressed mathematically as $\frac{\text{sine X}}{X}$ or (sine X)/X] to the pulse characteristics already named is of great importance because it reveals the bandwidth requirements for handling rectangular pulses.

1. How the (sine X)/X Curve is Plotted

The general appearance of the curve is similar to that of a sinusoidal wave whose peak amplitude is gradually decreasing. This curve may be plotted in a manner very similar to that used when the student undertook to plot a sinusoidal wave using an assigned peak voltage and a table of sines. It is not intended that the student should plot a (sine X)/X Curve but only that he should gain a thorough understanding of what one is. This may be accomplished best by showing how such a curve is plotted.

(a) How to Express "X". For the immediate demonstration let it be assumed that only 360° of the curve need be plotted and that one point will be plotted for each 30° of this total angle. To solve the relationship, (sine X)/X the angle represented by "X" must be expressed in radians. The conversion from degrees to radians is made in accordance with the procedure that was taught in Waveform Analysis. The desired angles are given below:

$$0^{\circ} = 0 \times \frac{2\pi}{360} = 0 \text{ radians}$$

$$30^{\circ} = 30 \times \frac{2\pi}{360} = \frac{\pi}{6} \text{radians}$$

$$60^{\circ} = 60 \times \frac{2\pi}{360} = \frac{\pi}{3} \text{ radians}$$

similarly,

$$90^{\circ} = \frac{\pi}{2} \text{ radians}$$

$$120^{\circ} = \frac{2\pi}{3} \text{ radians}$$

$150^{\circ} = \frac{5\pi}{6}$ radians

$180^{\circ} = \pi$ radians

$210^{\circ} = \frac{7\pi}{6}$ radians

$240^{\circ} = \frac{4\pi}{3}$ radians

$270^{\circ} = \frac{3\pi}{2}$ radians

$300^{\circ} = \frac{5\pi}{3}$ radians

$330^{\circ} = \frac{11\pi}{6}$ radians

$360^{\circ} = 2\pi$ radians

(b) Obtaining sine X. A table of sines was given in Waveform Analysis. The table provides the sines of angles from 0° to 360° in 30° increments just as is needed in the present problem. For convenience the table is repeated here.

Sine 0° = 0

Sine 30° = .5

Sine 60° = .866

Sine 90° = 1.

Sine 120° = .866

Sine 150° = .5

Sine 180° = 0.

Sine 210° = -.5

Sine 240° = -.866

Sine 270° = -1.

Sine 300° = -.866

Sine 330° = -.5

Sine 360° = 0

(c) Solving the Relationship (sine X)/X. All the values needed for solving the relationship (sine X)/X have been obtained. All that remains to be done is to take the successive values of the sine X

DOCID: 3726612

and divide them by the corresponding value of X expressed in radians. The process is as simple as all that--just the sine of the angle, X, divided by the angle, X. Note the results below:

When $x = 0^{o}$, (sine X)/X = (see explanation below)

When $x = 30^{o}$, $(\text{sine X})/X = \frac{\sin \pi/6}{\pi/6} = \frac{.5}{.5236} = .954$

When $x = 60^{o}$, $(\text{sine X})/X = \frac{\sin \pi/3}{\pi/3} = \frac{.866}{1.047} = .828$

When $x = 90^{o}$, (sine X)/X = .636

When $x = 120^{o}$, (sine X)/X = .414

When $x = 150^{o}$, (sine X)/X = .191

When $x = 180^{o}$, (sine X)/X = 0.

When $x = 210^{o}$, (sine X)/X = -.136

When $x = 240^{o}$, (sine X)/X = -.206

When $x = 270^{o}$, (sine X)/X = -.212

When $x = 300^{o}$, (sine X)/X = -.165

When $x = 330^{o}$, (sine X)/X = -.0868

When $x = 360^{o}$, (sine X)/X = 0

The one value needed but not given in the tabulation above is the value of (sine X)/X when X is equal to zero. This one value is obtained by mathematics of a higher level than that included in this study; however, the student is asked to accept the fact that (sine X)/X = 1 when X = 0.

(d) Facts Concerning the Curve. In Figure 26 the calculated values of (sine X)/X have been plotted on graph paper and the points smoothly joined so as to produce the (sine X)/X curve. In order to indicate more clearly the general character of its shape, the curve has been extended beyond the 360° for which calculations were made. In reality the (sine X)/X curve has no ending but continues without limit, its amplitude progressively approaching (but never actually reaching) zero. The significance of the curve so far as pulse spectrum bandwidth is concerned is that it provides an imaginary envelope or boundary that defines the peak amplitude of the fundamental component and each of its harmonics in any

DOCID: 3726612

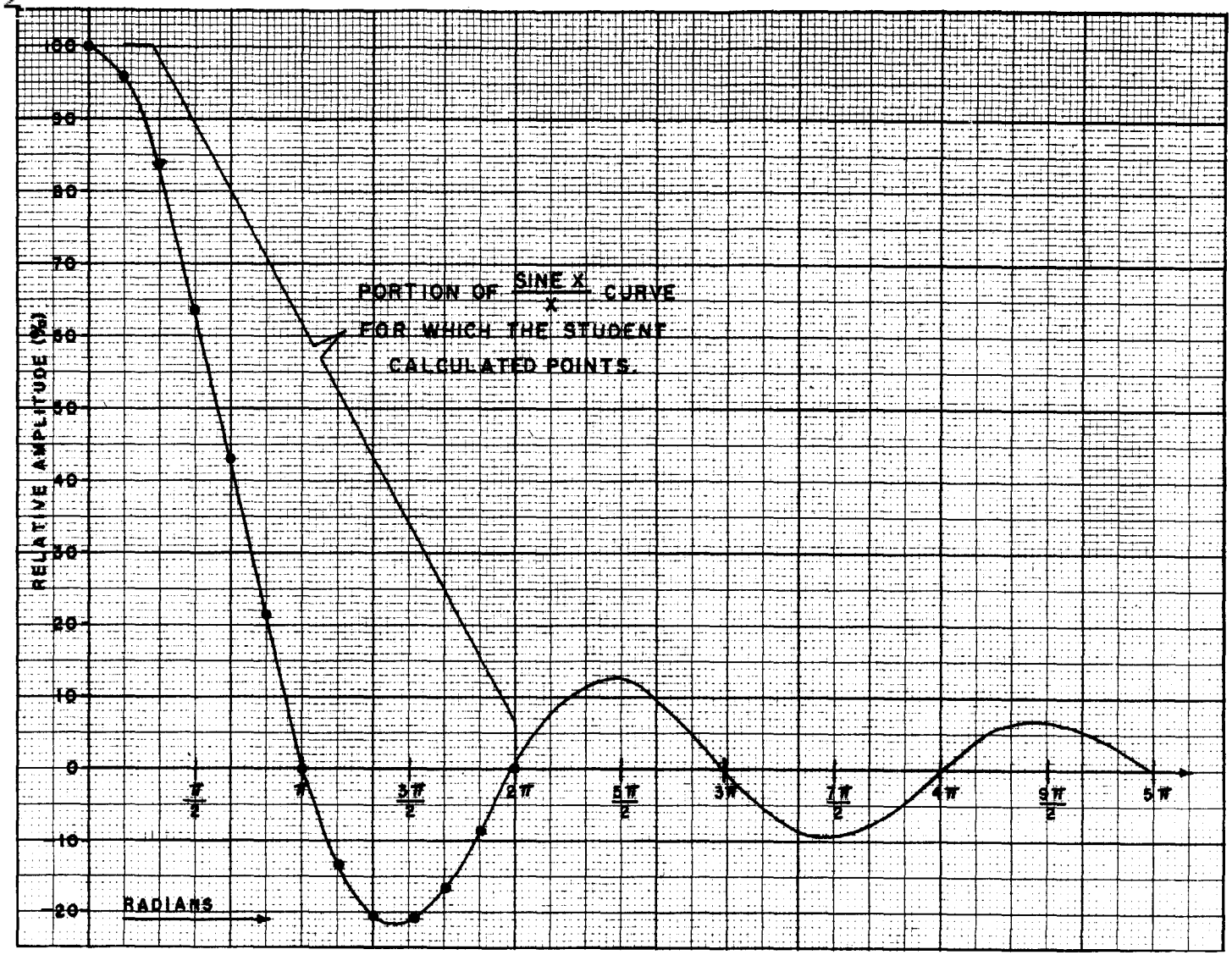


FIGURE 26. PLOT OF (SINE X)/X.

DOCID: 3726612

rectangular wave. For this reason the absolute height of each of the alternations (called lobes) of the curve is of more importance to us than their polarity (which is indicative of certain phase relationships). With this in mind, the (sine X)/X curve has been modified in Figure 27 to show amplitude only. Figure 27 is indicative of the form in which the curve will be drawn for the balance of this text. The points where the (sine X)/X curve originally crossed the horizontal reference line (Figure 26) are now points where the curve meets this reference line (Figure 27). These points will hereafter be referred to as (sine X)/X dropouts. These dropouts are all equally spaced regardless of the amplitude of their associated lobes. This makes all the lobes the same width at the dropouts points.

2. How the (sine X)/X Curve is Interpreted

The curve has been developed and the student has become acquainted with its appearance. Now it is necessary to see how it is applied in interpreting pulse spectrum bandwidth. (Note that the expression "pulse spectrum bandwidth" is used instead of pulse bandwidth. This is done because, after all, it is the spectrum of components that require bandwidth. The pulse is only a result of these.) This requires that the curve itself, and the reference lines along which it is drawn all be interpreted in terms of amplitude and frequency. The procedure of interpretation is not involved and if the student follows closely no difficulty should be encountered.

(a) The Vertical Reference Line. The length of the vertical reference line from its origin to the point where it is intercepted by the curve represents double the peak amplitude of the pulse multiplied by the duty cycle. This, expressed mathematically is $2A\frac{W}{P}$. (In the case of the true transient this relationship has no use whatever. This will be developed later in the lesson.) The amplitude of all sinusoidal components of the wave must be relative to this value and none can equal nor exceed it. Dc values (when indicated) must be superimposed on this axis. Since all other amplitudes are relative to the length of this reference, the line can be scaled in terms of the percentages of its full length. This is done in Figure 28.

(b) The Horizontal Reference Line. This reference line will become a frequency axis. Its length starts at zero frequency (dc) where it intersects with the vertical reference line, and extends

DOCID: 3726612

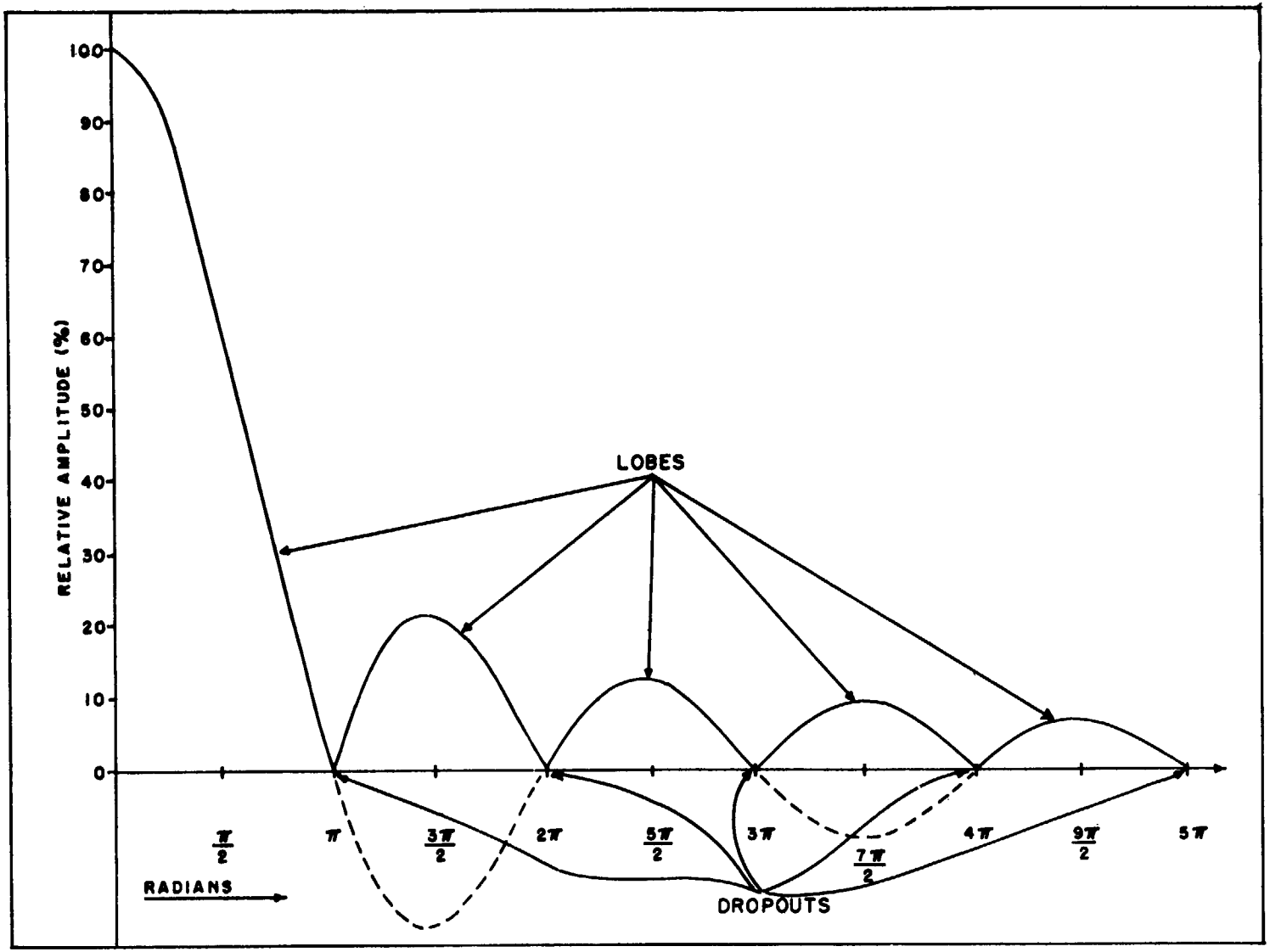


FIGURE 27. MODIFIED (SINE X)/X CURVE.

DOCID: 3726612

out in frequency to infinity. We will never take it out beyond the fifth (sine X)/X dropout which occurs at 5 π radians of the curve shown in Figure 28.

(c) Significance of the Dropouts. Each of the (sine X)/X dropouts will represent some frequency along the frequency axis. In fact the dropouts will provide the relative scale for the frequency axis. The dropout corresponding to π radians will represent a frequency equal to the reciprocal of the pulse duration ($\frac{1}{W}$) of any rectangular pulse. (A note of caution must be included here, namely, that when evaluating (sine X)/X dropouts the shortest element, regardless of its polarity, must be considered as the pulse.) The value of the dropout corresponding to 2 π radians is simply twice the frequency of the first (i.e. $\frac{2}{W}$); the drop out corresponding to 3π radians is three times the value of the first (i.e. $\frac{3}{W}$) and so on. Provision is made for labeling the dropouts for a particular rectangular wave by supplying brackets at the appropriate points in Figure 28.

3. Practical Application of the (sine X)/X Curve

The method of interpreting the curve in pulse spectrum analysis is best understood after working examples for a number of rectangular pulses. The following illustrations should help in this matter.

(a) Example 1. The pulse duration is 1 microsecond, and the pulse period is 2 microseconds (refer to Figure 29). The duty cycle is necessarily 50%. The first thing to do is calculate the value of the dropouts. The first dropout is given by

$$\frac{1}{W} = \frac{1}{1 \times 10^{-6} \text{ seconds}} = 1 \text{ megacycle.}$$

This value has been placed in the appropriate brackets in Figure 29. The second dropout is given by

$$\frac{2}{W} = \frac{2}{1 \times 10^{-6}} = 2 \text{ megacycles.}$$

This value has also been placed in the appropriate brackets. The student should now calculate the values for the three remaining dropouts and write the values in the proper brackets in Figure 29. With this done, the (sine X)/X curve is calibrated with the appropriate frequency scale for the wave specified.

DOCID: 3726612

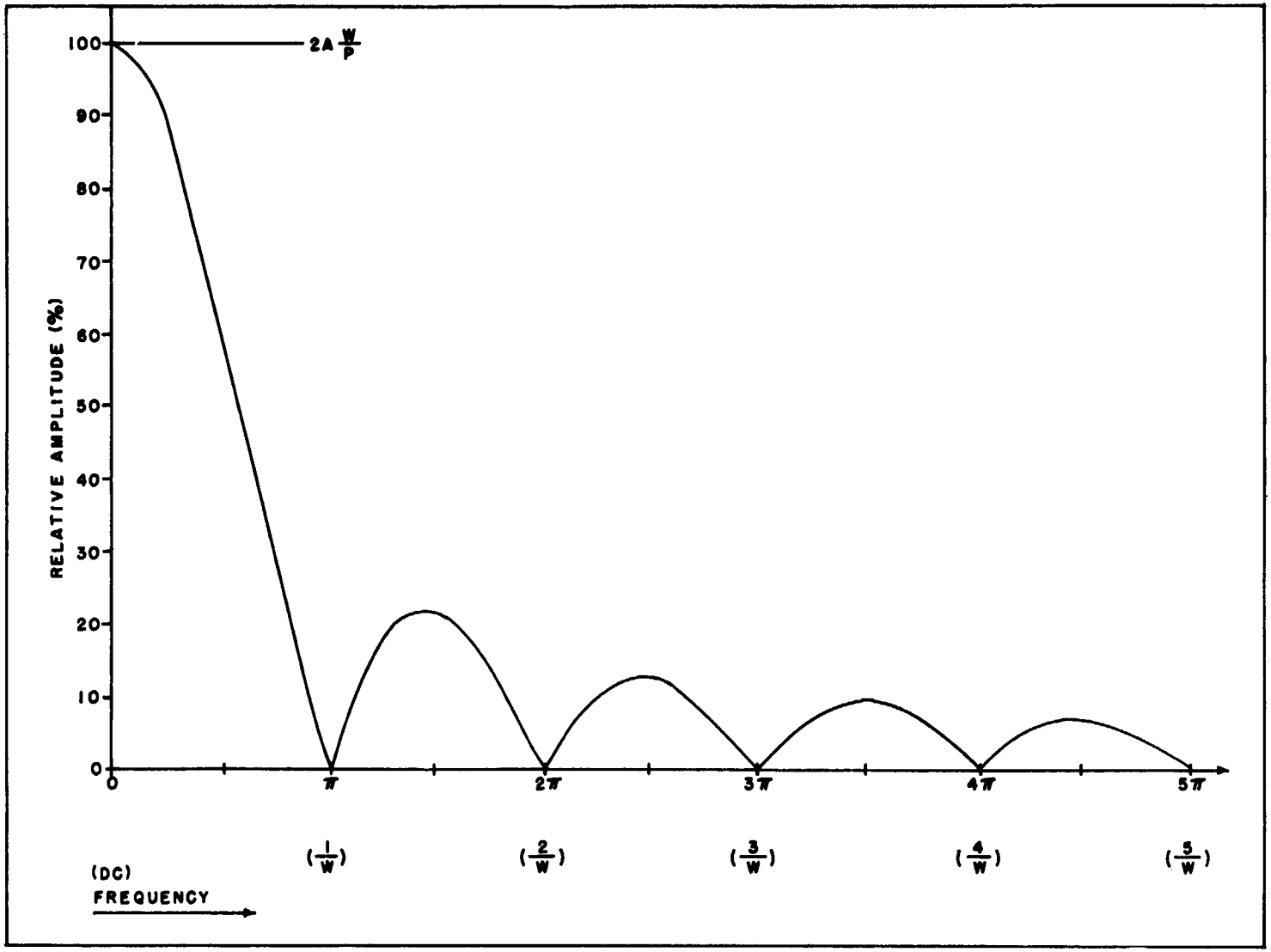


FIGURE 28. INTERPRETATION OF THE CURVE.

DOCID: 3726612

The second thing to do is to calculate the value of the fundamental component of the sinewave. This is given by

$$f_1 = \frac{1}{P} = \frac{1}{2 \times 10^{-6}} = .5 \text{ mc}$$

Notice that this frequency is one-half as great as 1 megacycle, therefore it is plotted along the frequency axis half-way between zero (or dc) and 1 megacycle in Figure 29. It appears as a vertical line whose amplitude is defined by the point where it joins the (sine X)/X curve.

The second harmonic should be calculated and plotted next. Its value is given by

$$f_2 = \frac{2}{P} = \frac{2}{2 \times 10^{-6}} = 1 \text{ mc}$$

or simply double the value of the fundamental. It falls exactly on the first dropout. The (sine X)/X curve has zero amplitude at this point, therefore the amplitude of this harmonic is ZERO. The third harmonic is next in order. Its value is given by

$$f_3 = \frac{3}{P} = \frac{3}{2 \times 10^{-6}} = 1.5 \text{ mc}$$

or simply three times the fundamental. This frequency is midway between 1 megacycle and two megacycles, therefore it is plotted half-way between the first and second dropouts in Figure 29. As before, its amplitude is defined by the height of the (sine X)/X curve at this particular point.

The fourth harmonic should be found by the method already prescribed. It will be noted that its frequency will be exactly coincident with the second dropout, which means that its amplitude is zero.

It is left to the student (under the direction of the instructor, if necessary) to compute harmonics 5 through 10 and plot them correctly in Figure 29.

Summarizing the results, the vertical lines representing the sinusoidal components of the pulse train are called spectral lines. Since these components are harmonically related, the spectral lines are separated by multiples of the PRF or multiples of $\frac{1}{P}$. The dropouts are also harmonically related to each other but are separated by multiples of $\frac{1}{W}$. Also it is important to observe here that for a squarewave all even harmonics are coincident with the points where the (sine X)/X curve falls to zero and thus they "drop out." This latter fact bears out a previous claim that squarewaves consist

DOCID: 3726612

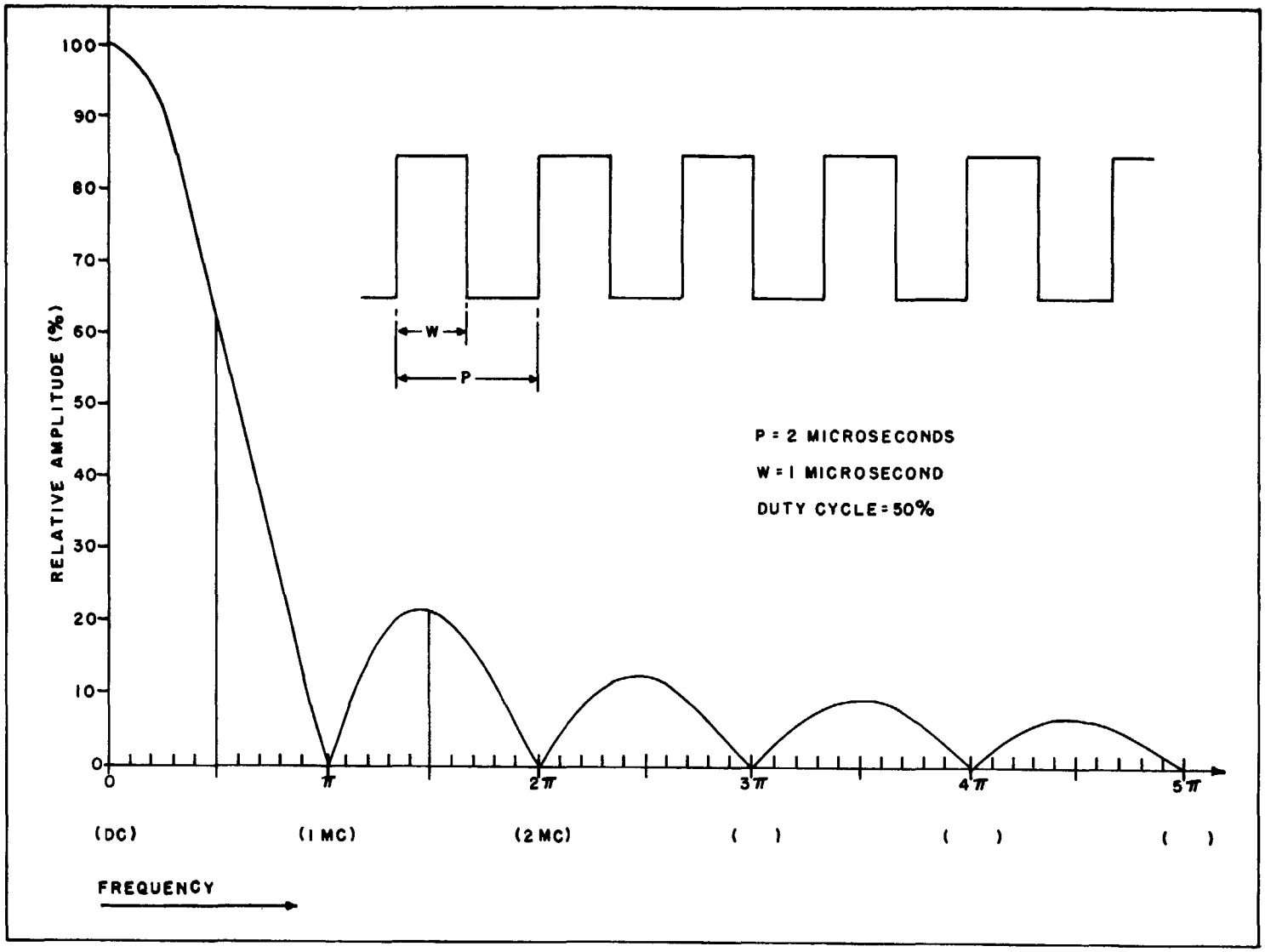


FIGURE 29. EXAMPLE 1, A SQUARE WAVE.

DOCID: 3726612

of only odd harmonics. Figures 30 and 31, reprinted from Waveform Analysis, should refresh the student's memory on these points. Notice that Figure 31B shows the same spectrum as Figure 29 with the exception that the (sine X)/X curve is not drawn in.

(b) Example 2. Keeping the period the same as it was in Example 1, let the pulse duration be decreased to .5 microsecond. This gives a duty cycle of 25%. Obviously the wave is still rectangular but it is no longer a squarewave. The wave is shown in Figure 32. As with Example 1, the value of the dropout should be computed first. This provides a frequency scale with which to work. The first and second dropouts occur at 2 megacycles and 4 megacycles and are so labeled on the curve of Figure 32. It will be left to the student to compute the others and properly label the appropriate points.

The fundamental frequency should be calculated next. It, of course, will not be different from the value in the preceding example since the period of the wave has not been changed. This value is .5 megacycles. The frequencies of the harmonics will be unchanged also. The change in pulse duration has created a very important difference, however, that appears as new positions for the spectral lines with respect to the dropouts. In this example the first spectral component is located only one-quarter of the distance to the first dropout because .5 megacycles is only one-fourth of 2 megacycles. The distance between harmonics is equal to the fundamental frequency and so this places the second and third harmonics within the first lobe. The fourth harmonic drops out. The student should draw in the remainder of the spectral lines and label each one as to its frequency.

Since this wave is not a squarewave not all of the even harmonics are coincident with dropouts; hence it is observed that rectangular waves that are not square can have even multiples of the fundamental frequency.

(c) Example 3. Figure 33 shows a train of pulses where an equipment is keyed on for a long time and is off for a relatively short time. The pulse period is 2 microseconds just as in the two previous examples but the pulse duration is 1.5 microseconds. For purposes of computing the duty cycle, the pulse is still the "ON-TIME" just as it has been before (this gives a duty cycle of $\frac{1.5}{2}$ = 75%) but for the purpose of computing the spectrum, THE SHORTEST ELEMENT MUST ALWAYS BE CONSIDERED TO BE THE PULSE.

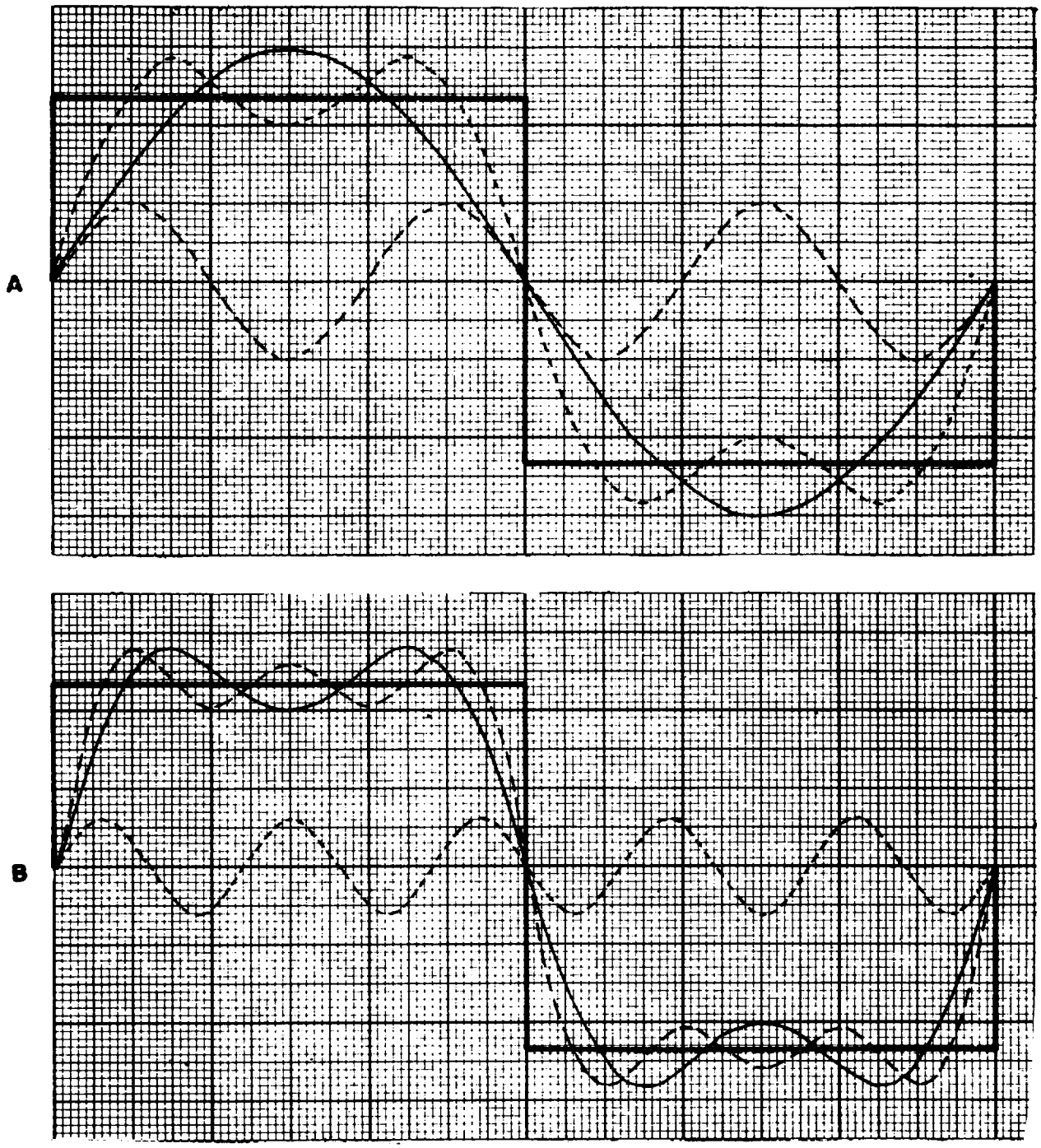


FIGURE 30. COMPONENTS OF A SQUAREWAVE

DOCID: 3726612

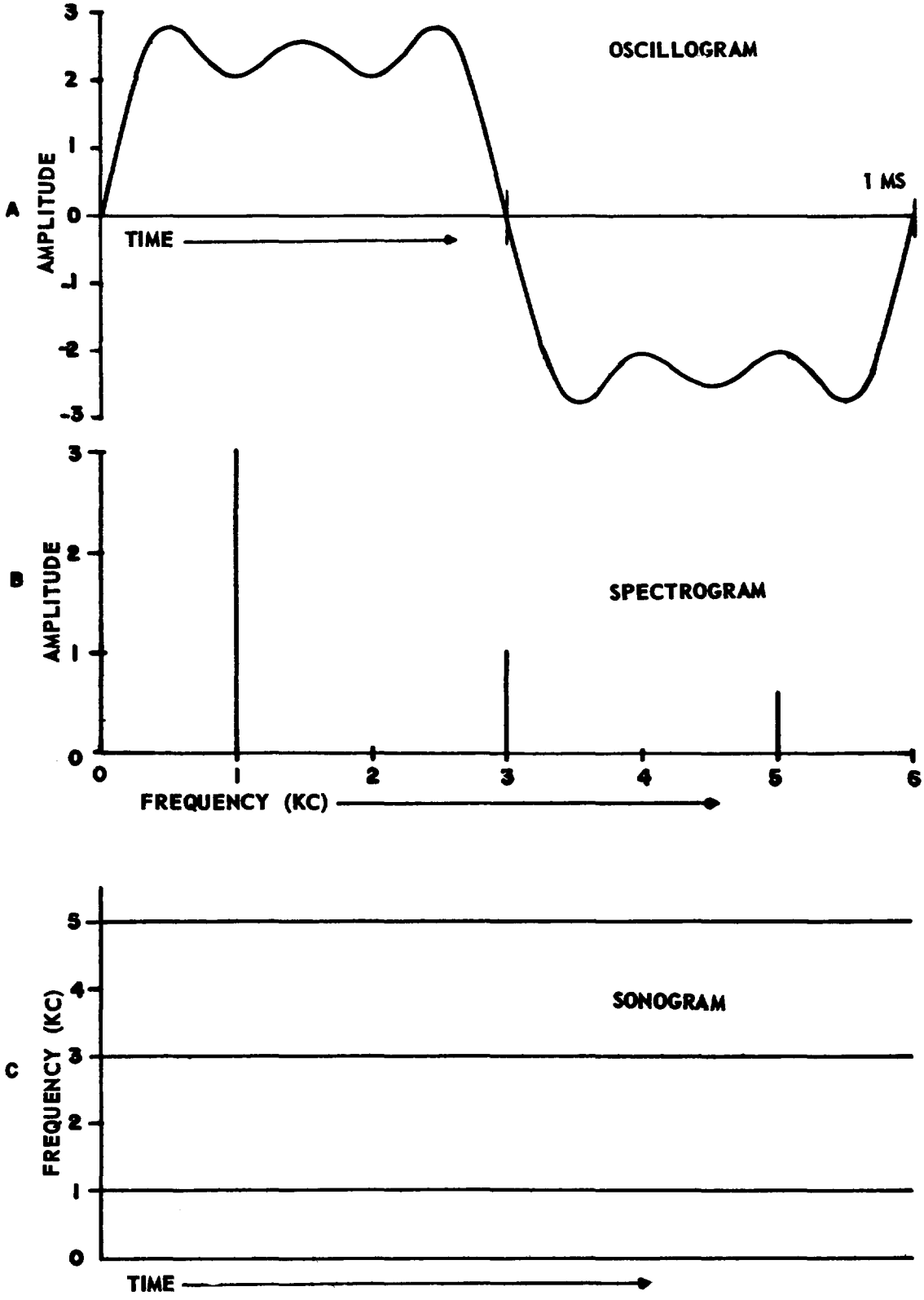


FIGURE 31. THREE VIEWS OF A SQUAREWAVE

DOCID: 3726612

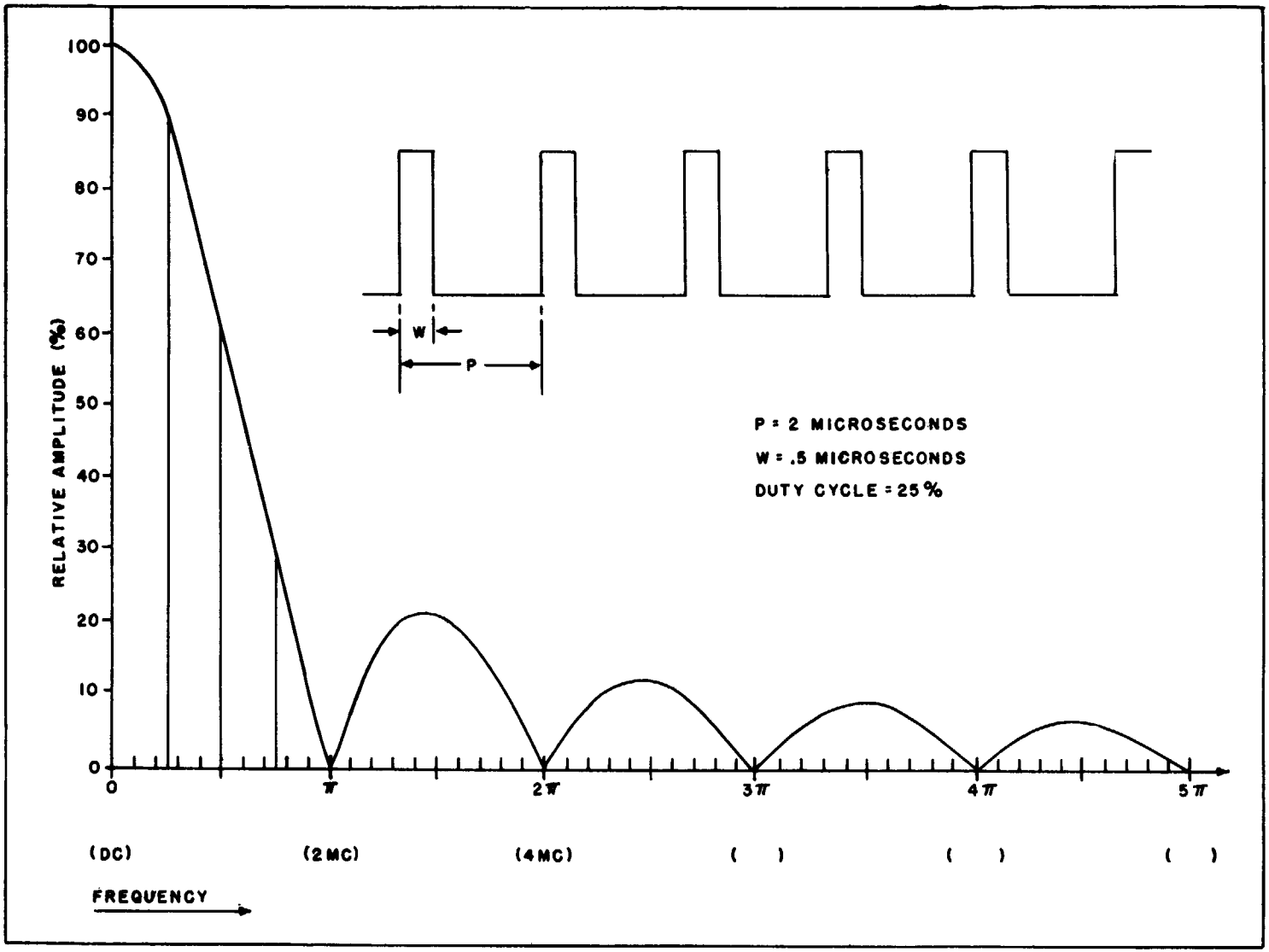


FIGURE 32. EXAMPLE 2, NARROW PULSES.

DOCID: 3726612

The student should fill out the spectrum for Figure 33 and note that there is no difference between this spectrum and Figure 32. Dropouts are still 2 megacycles apart, and the spectral lines are still .5 megacycles apart. The only change is in the Waveform itself in that it has been inverted so that ON and OFF times have exchanged places.

(d) Example 4. Narrowing the pulse duration, as we have seen, widens the spectrum bandwidth and widening the pulse duration narrows the required spectrum bandwidth provided the duty cycle is not increased beyond 50%, in which case the bandwidth again increases since the dropouts are based on the shortest element. This present example will serve to illustrate the effects of changing the pulse period.

Consider again the pulse train of Figure 29. If the same pulse duration is maintained but the pulses are spaced 4 times as far apart, the waveform of Figure 34 results. This gives a pulse period of 5 microseconds, a pulse duration of 1 microsecond and a 20% duty cycle.

In this example it will be left for the student to compute and label all dropouts, and to compute, plot and label all spectral lines. The outstanding fact to note in this example is that the dropouts all occur at the same frequencies as they did in example 1--this is because the pulse duration has not undergone any change. The spectral lines, on the other hand, are much more numerous for the same bandwidth and much closer together.

(e) Example 5. Figure 35 carries the idea of the last example still further, namely that if the pulse period is increased more and more, the harmonics represented by spectral lines get closer and closer. Here the pulse duration has not been changed--it remains at 1 microsecond--but the pulse period is double the value it was in the last example and ten times as great as it was in example 1. There are now twice as many frequencies present in the same bandwidth as there were in the previous example. The student should draw in and label all values.

(f) Example 6. In Example 5 there were 10 frequencies contained in the first lobe (although the 10th harmonic dropped out). Considering bandwidth up to the 5th dropout, space is made for 50 harmonics, although harmonics 10, 20, 30, 40, and 50 all dropped out.

DOCID: 3726612

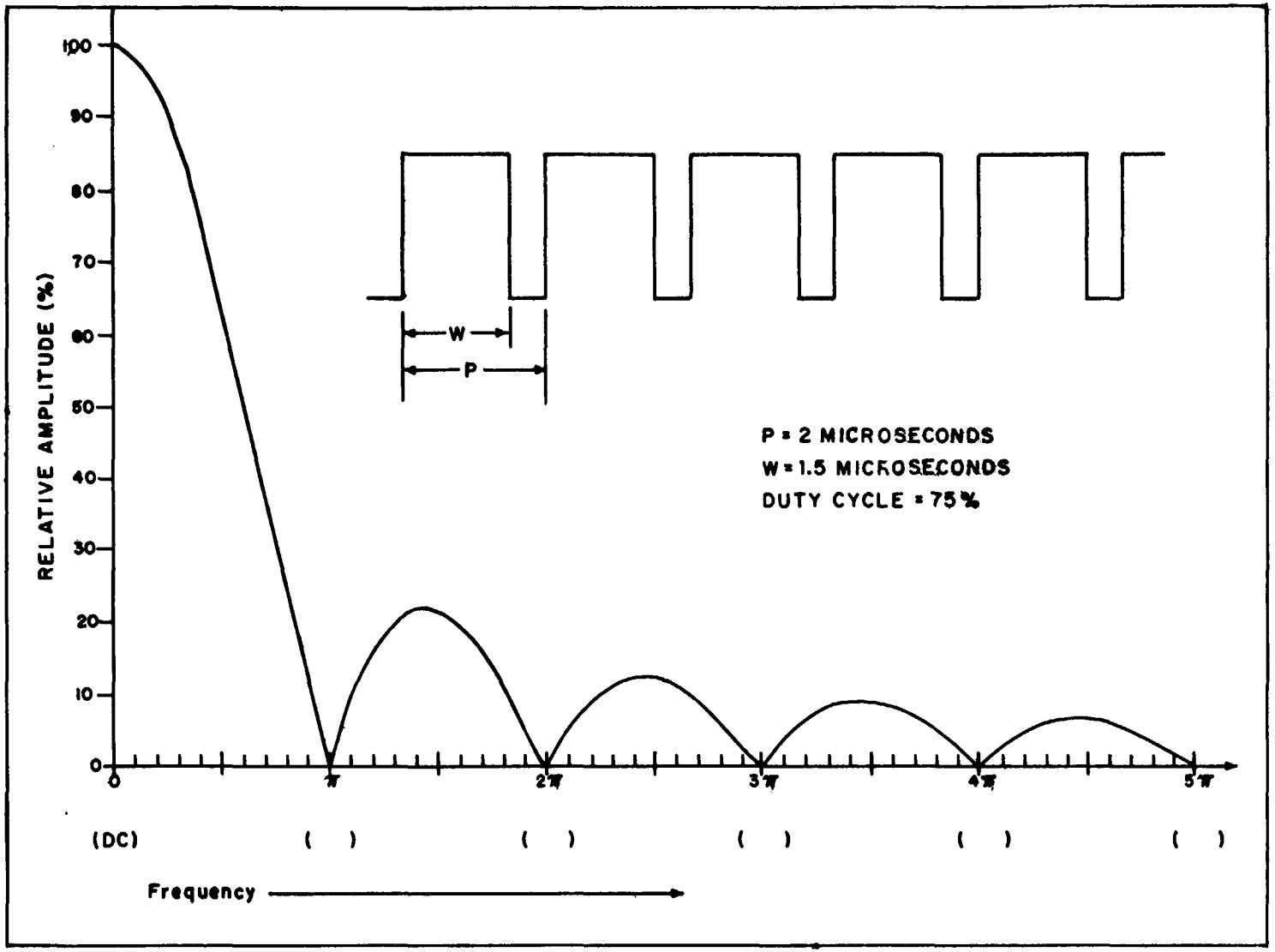


FIGURE 33. EXAMPLE 3, THE SHORTEST ELEMENT DETERMINES SPECTRUM

DOCID: 3726612

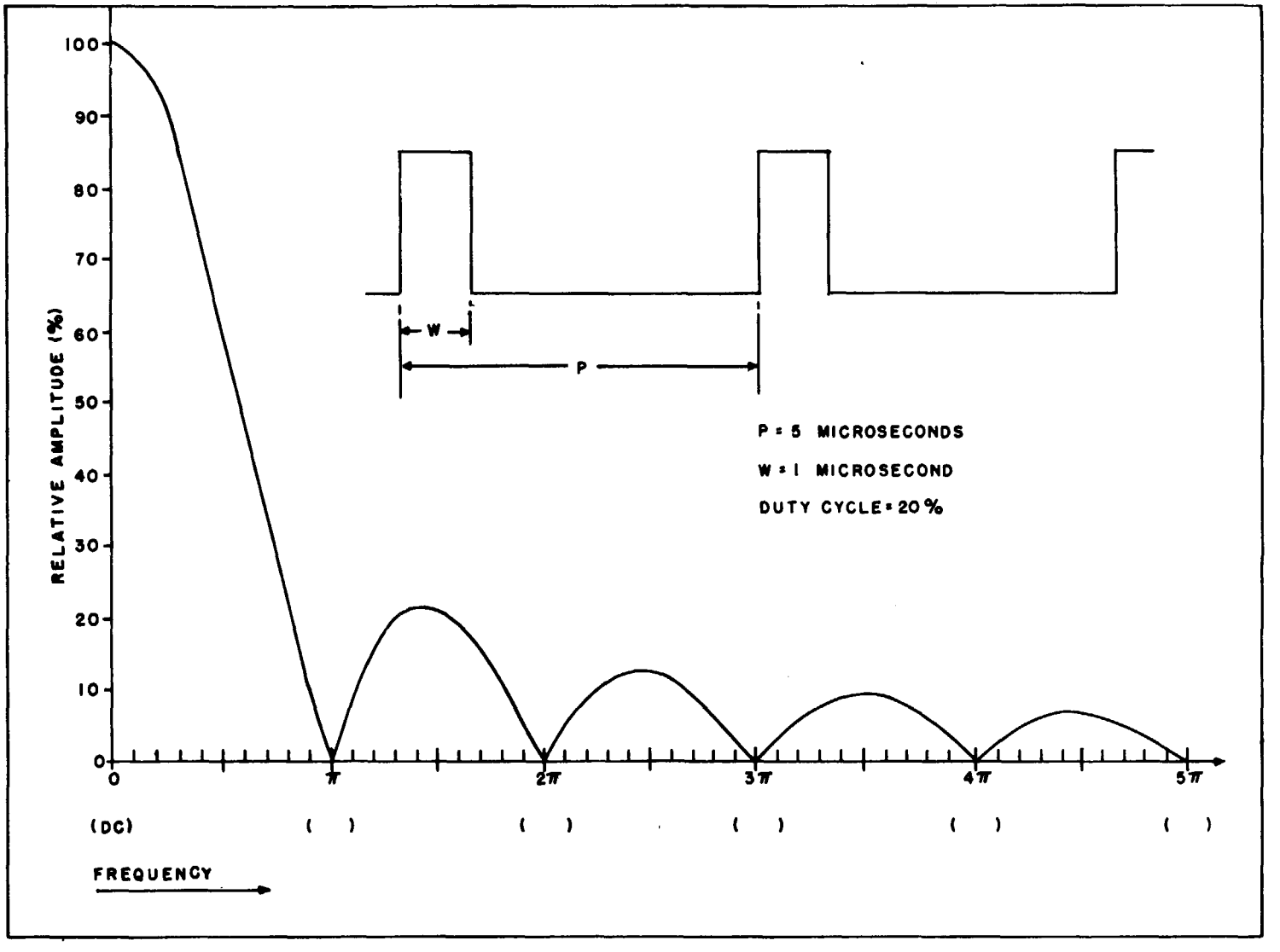


FIGURE 34. EXAMPLE 4, PULSE PERIOD INCREASE TO 5 μ SEC.

DOCID: 3726612

Suppose that the period were increased again by a factor of 10. Five dropouts would then include not 50 but 500 harmonics. Already there would be too many for the student to draw. But, to go on, if the pulse period were to become still another ten times longer (only 1 millisecond at that) there would be 5000 harmonics (spectral lines) contained within the first five dropouts. For the scale used, there would not be discernable spacing between the pencil lines used to represent these harmonics. The scene is now set for the situation of Figure 36 where a pulse is not repeated (i.e., the period is infinite). The average dc value of this wave is zero, there are no sinewave components, it has no fundamental frequency nor any harmonics, its duty cycle is zero, yet it has an energy distribution throughout the entire frequency spectrum, that is, there is some very small amount of energy to be found at all frequencies except, of course, at the dropouts which are based on pulse duration and not on pulse period. All frequencies (essentially) are components of this wave. Except in this last instance all examples used to illustrate the relationship of the (sine X)/X curve to pulse spectrum bandwidth considered only periodic rectangular pulses but this wave is important because it it not periodic. It is called a transient. The resultant spectrum may be considered typical of any rectangular pulse that either is not repeated, or if repeated at all, the repetition is not at any fixed regular interval. Transients or aperiodic waves are always encountered in communications because lack of periodicity is characteristic of any signal intended to convey intelligence. This is because intelligence, by its very nature, cannot be periodic.

B. Recapitulation of Pulse Duration & Pulse Period Effects

The examples under the last topic bring to light several facts, which, although they have already been mentioned cannot be over-emphasized. These facts will be recapitulated here.

1. The Effect of Varying Pulse Duration

When Examples 1, 2, and 3 are compared notice that an increase of pulse duration decreases the bandwidth of the lobes and a decrease of the pulse duration increases the bandwidth of the lobes. This effect exists because the bandwidth of the lobes (bandwidth between dropouts) and the pulse duration are reciprocals of each other. Described in different words, pulse duration and bandwidth vary inversely with respect to each other. Since it is the shortening of pulse duration

DOCID: 3726612

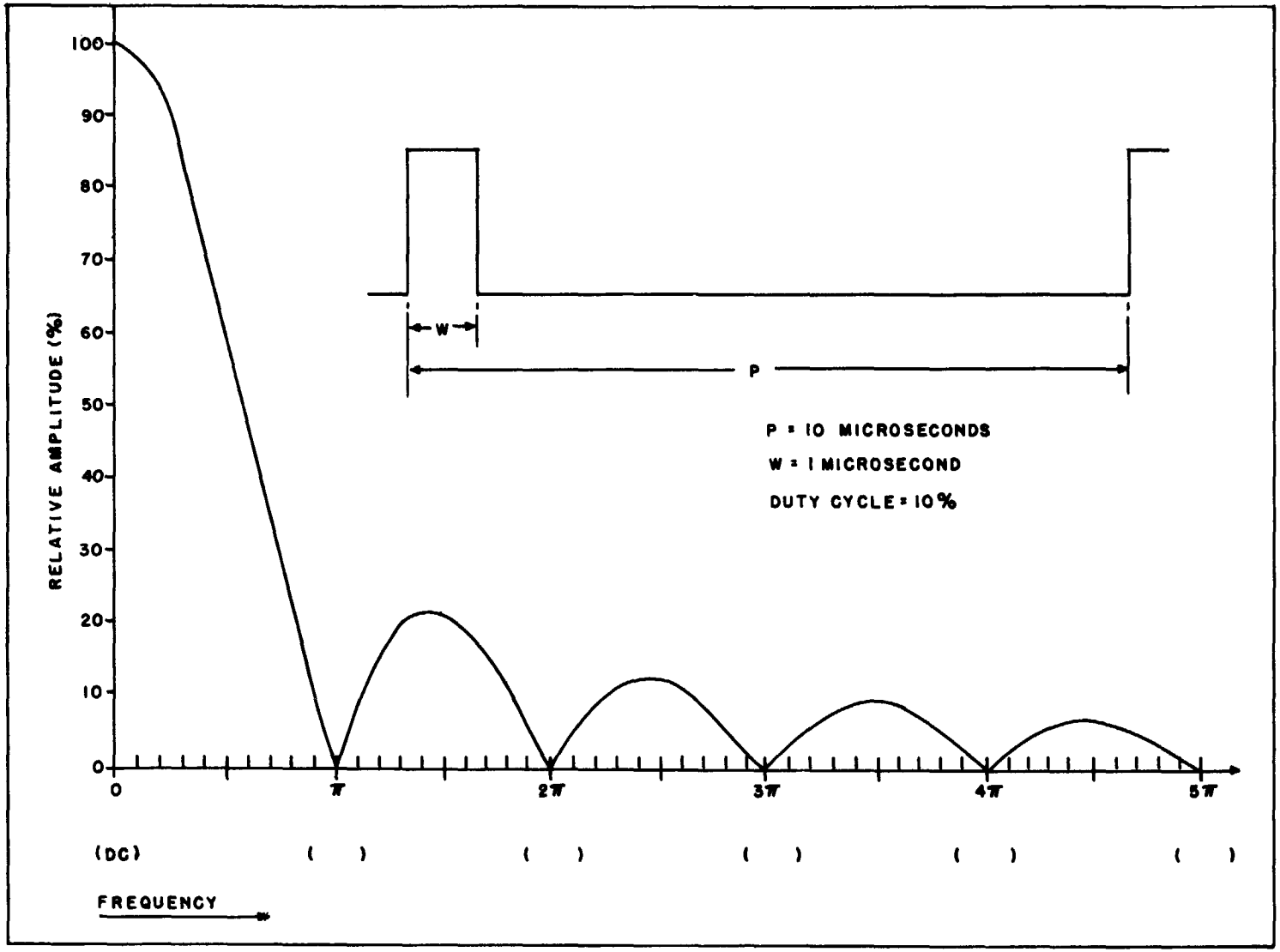


FIGURE 35. EXAMPLE 5, PULSE PERIOD INCREASED TO 10μ SEC.

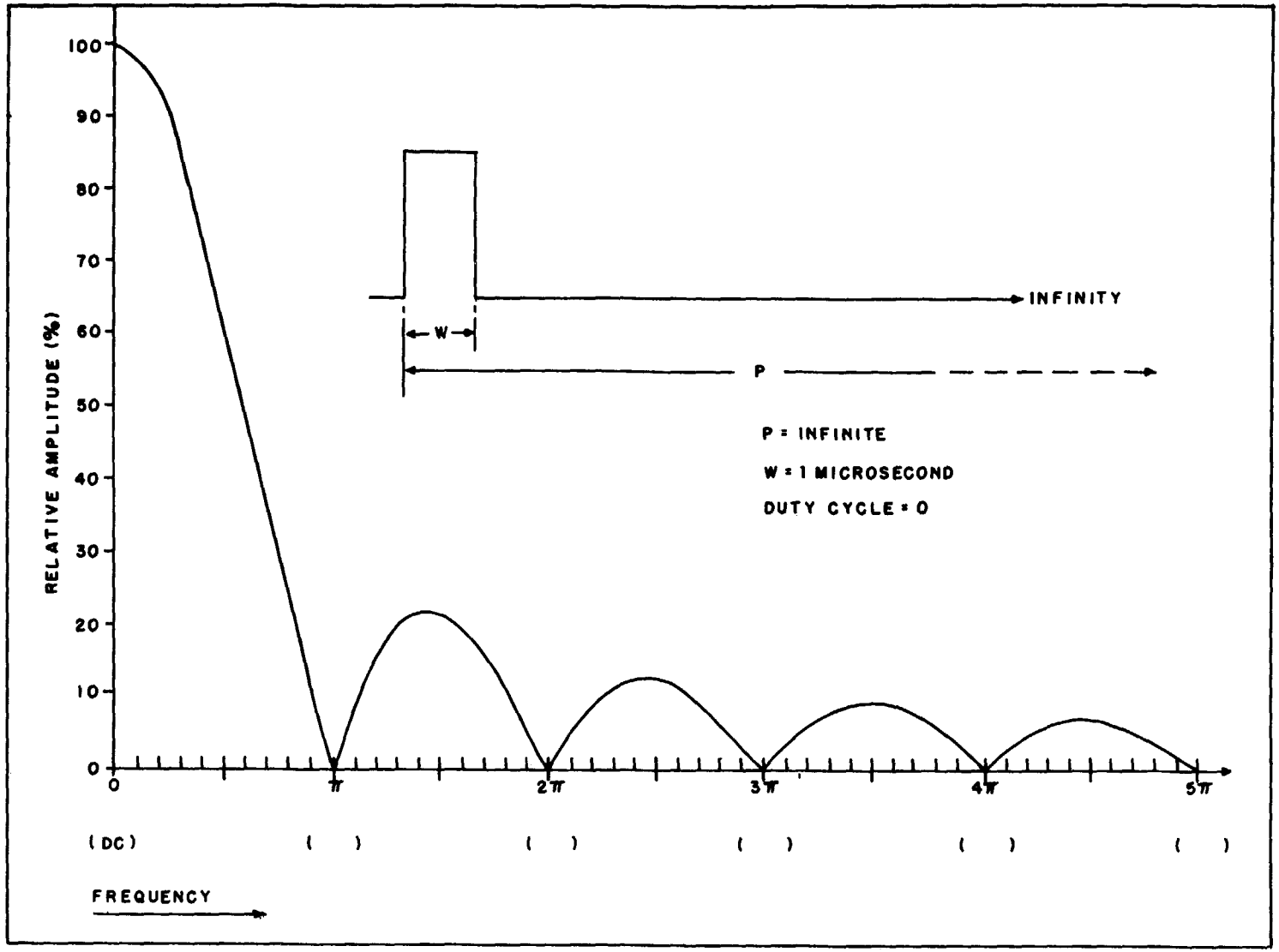


FIGURE 36. EXAMPLE 6, PULSE PERIOD INCREASED INFINITELY.

DOCID: 3726612

that demands the widening of bandwidth, the shortest keying element must always be regarded as the pulse in bandwidth calculations. It matters not whether this "shortest" element represents power applied or power removed. It can be summarized, then, that irrespective of pulse period short pulse durations require wide bandwidth and long pulse durations require narrow bandwidths.

2. The Effect of Varying Pulse Period

When Examples 4, 5, and 6 are compared to Example 1, notice that variations in the pulse period do not cause any changes in the bandwidth of the lobes so long as the pulse duration is unchanged. For a fixed pulse duration two extreme conditions can be described, the one occurring when the wave is an isolated transient and the other occurring when the wave is a periodic squarewave. The isolated transient will have an infinite number of frequencies under each lobe whereas the periodic squarewave will have one harmonic of its fundamental component under each lobe. All other conditions exist somewhere in between these two extremes. Pulse period, then, simply regulates the number of frequencies present within each lobe of the (sine X)/X curve. The separation of harmonics and the pulse period are reciprocals of each other, that is, the separation of harmonics and the pulse period vary inversely with respect to each other. Summarizing, short pulse periods place the harmonics far apart and long pulse periods bring the harmonics close together.

3. Pulse Duration, Pulse Period and Keying Rate

When the keying rate is increased the period and duration of the pulse both decrease. The separation of (sine X)/X dropouts (lobe bandwidth) and the separation of harmonics become greater. Therefore the increased keying rate extends the minimum bandwidth requirement for handling the pulsed information. If the bandwidth of an equipment were insufficient to accommodate the wider signal bandwidth the signal would be distorted to an extent that may be anywhere from slight to beyond recognition. Equipment bandwidth, therefore, limits the keying rate that can be handled properly.

C. Bandwidth Requirements

So far as the significant bandwidth for a transient is concerned, the most important lobes of the envelope are those containing the most energy, namely, the larger ones. When considering the pulse spectrum bandwidth to just the fifth

DOCID: 3726612

(sine X)/X dropout as has been donehere,roughly 60% of the energy in the spectrum is contained in the first lobe, and another 20% is contained in the second lobe. The combined energy in the remaining three lobes totals only another 20% (approximately). Consequently it is not difficult to draw the obvious conclusion that the pulse spectrum bandwidth up to the second dropout is most significant since 80% of the total energy is represented here.

1. Squarewave Fidelity

Figure 37 compares two one-millisecond pulses, one of which is periodic and the other a transient. If the transmitting and receiving equipment were to allow passage of the pulse spectrum bandwidth up to the first dropout, about 60% of the energy in the spectrum of either pulse would be passed. This constitutes over one-half of the spectral energy which, of course, would provide a substantial signal even if all the rest of the spectrum were ignored, but just a substantial signal is not the ultimate requirement for minimum bandwidth.

The ultimate requirement for minimum bandwidth is determined by the amount of distortion that can be tolerated for a squarewave having the smallest pulse duration that will be encountered in a given system. (The distortion considered in this case is that which results from the loss of harmonic components.) For the transient, a pulse spectrum bandwidth up to the first dropout will give a very good representation of a rectangular pulse but for the squarewave only the fundamental frequency--a sinewave--will appear, the higher harmonics occurring too far up in the spectrum to be passed.

If the bandpass of the equipment were to be extended to admit the spectrum up to the second dropout the total energy passed would be about 80% which would substantially reduce the distortion of the transient and would allow enough of the squarewave spectrum to pass that it would appear as the resultant in Figure 30A. A still better approximation of the squarewave could be achieved if the bandpass were extended still further to the third dropout which would allow the fifth harmonic to pass without alternation. The appearance of the squarewave would then be that of the resultant shown in Figure 30B.

2. Minimum Fidelity

The previous observations based on Figure 37 illustrate that minimum bandwidth requirements must be based on that bandwidth which is required to pass a squarewave of some specified minimum

DOCID: 3726612

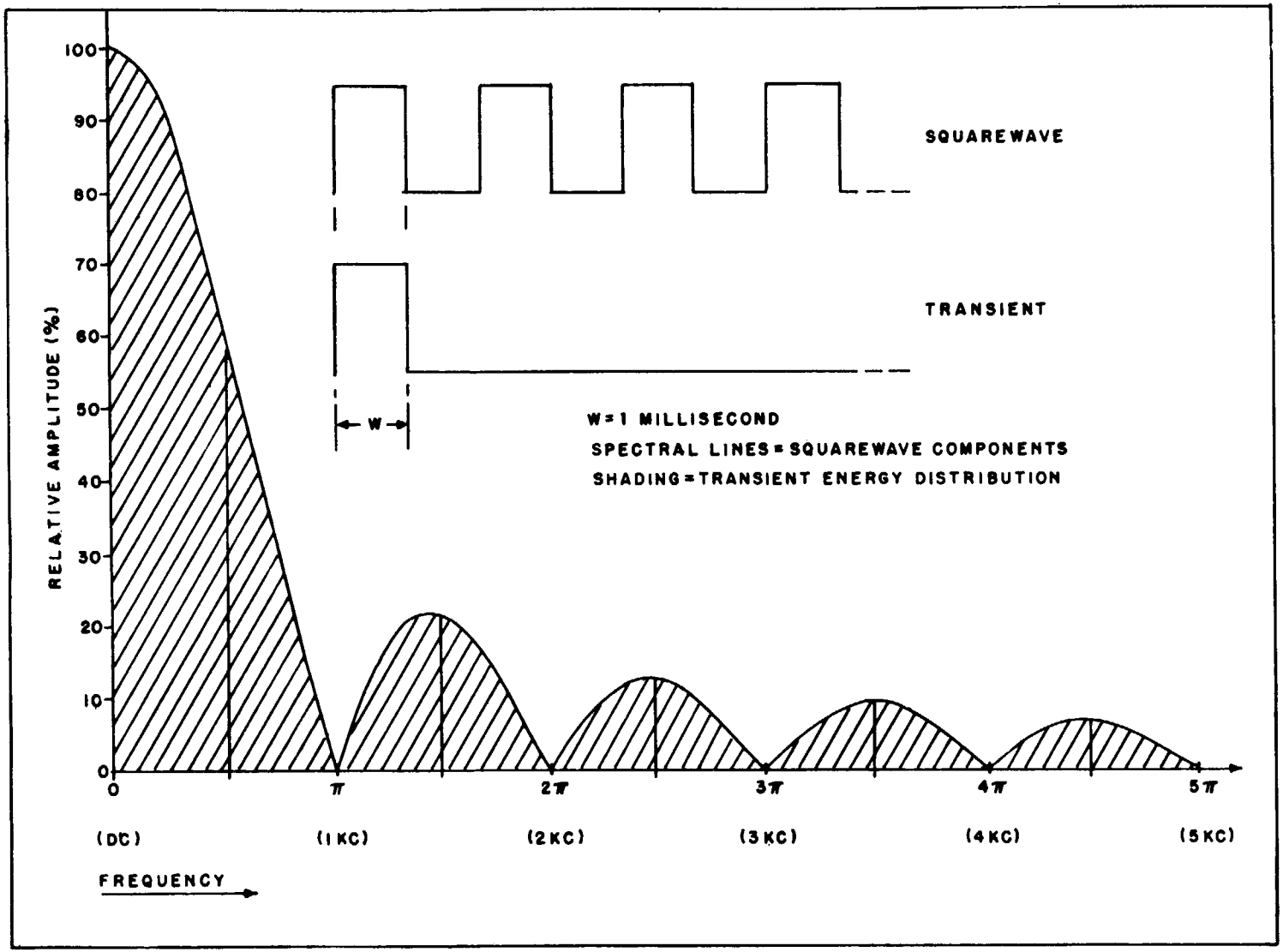


FIGURE 37. A COMPARISON OF PERIODIC AND TRANSIENT WAVE SPECTRA.

pulse duration without detrimental loss of fidelity. When this requirement is met, any rectangular pulse (whether periodic or transient) which has equal or greater pulse duration can be reproduced with equal or greater fidelity.

The question naturally arises, "What is the minimum acceptable fidelity?" This question cannot be given a precise answer. The answer in any particular instance depends upon the equipment in use. Equipments of some systems use a bandwidth equal to 3/4 the width of the first (sine X)/X lobe--this is just 1 1/2 times the fundamental frequency of a squarewave. Three-fourths of the width of the first (sine X)/X lobe is more than enough bandwidth to pass the fundamental sinewave, the additional width insuring that it will be passed without attenuation.

The use of such a narrow bandwidth as that just described is more the exception than the rule, and is, indeed, an extreme case. Different systems will have different bandwidth requirements. In general, pulse transmissions carried by landline are free from fading, and as a result need only to include the fundamental and third harmonics of a squarewave. Pulse transmissions sent by radio (appropriately modulating a carrier is to be understood), on the other hand are subject to fading, and to overcome this difficulty the fundamental, third, and fifth harmonics of a squarewave must be considered as the minimum intelligence bandwidth if the waveform is to be reproduced reasonably well.

Where squarewave pulse components to the third harmonic are sufficient a little extra tolerance in bandwidth is allowed which includes the second (sine X)/X dropout. This would be a bandwidth up to the fourth harmonic although this harmonic has no amplitude for a squarewave. The extra tolerance simply provides for the desired spectrum to pass unattenuated.

3. DC Components

Perhaps the easiest pulse to generate is that type which is formed by simply switching a circuit between two dc voltage levels. The switching may be back and forth between one positive and one negative level or between two levels of the same polarity, or yet simplest of all between OFF and ON (i.e., between zero and some other level that is either positive or negative with respect to zero).

DOCID: 3726612

Switching in any of the ways described produces a significant result. The very action of switching causes the voltage applied to the circuit to alternate. Thus, while switching is taking place the voltage applied to the circuit is converted from a dc (i.e., continuous) voltage to an ac (i.e., varying) voltage. The resultant wave, although no longer dc, may still have a dc component.

The dc component, if present, is the average value of the wave over its complete cycle. For example, a polar-keyed squarewave has equal positive and negative excursions of equal duration. The average value of this wave over any complete cycle is zero because in respect to time and amplitude there is as much of the wave above zero as there is below zero. The dc component is, therefore, zero also. On the other hand, a neutral-keyed squarewave is produced by keying a circuit ON and OFF. If the peak amplitude of such a wave were +10 volts, the average value of the wave would be +5 volts and therefore the dc component of the wave is +5 volts.

The presence or absence of a dc component determines the position of the wave with respect to the zero voltage axis but has no effect on the wave so far as its shape is concerned (See Figure 38). The waveshape is determined solely by the ac components and their relationship to one another in frequency, phase and amplitude. Obviously, removal of the dc component cannot in any way alter the mutual relationship of the ac components and therefore their part of the wave spectrum is unchanged. Since, by its very nature, intelligence is alternating, a dc component is not essential to the intelligence content of a wave; thus the dc component can be removed without any loss of intelligence.

Figure 38 illustrates the fact that it is not necessary to include the dc component in the pulse spectrum bandwidth by comparing polar- and neutral-keyed squarewaves. In the polar-keyed wave the dc component is removed or suppressed, whereas in the neutral-keyed wave the dc component is retained. In both instances the waveshape is unaffected and the spectrum of ac components is unaltered. Another point of interest in this figure is that the amplitude of the dc component, even when present, is not regulated by the (sine X)/X curve as the amplitudes of the ac components are.

DOCID: 3726612

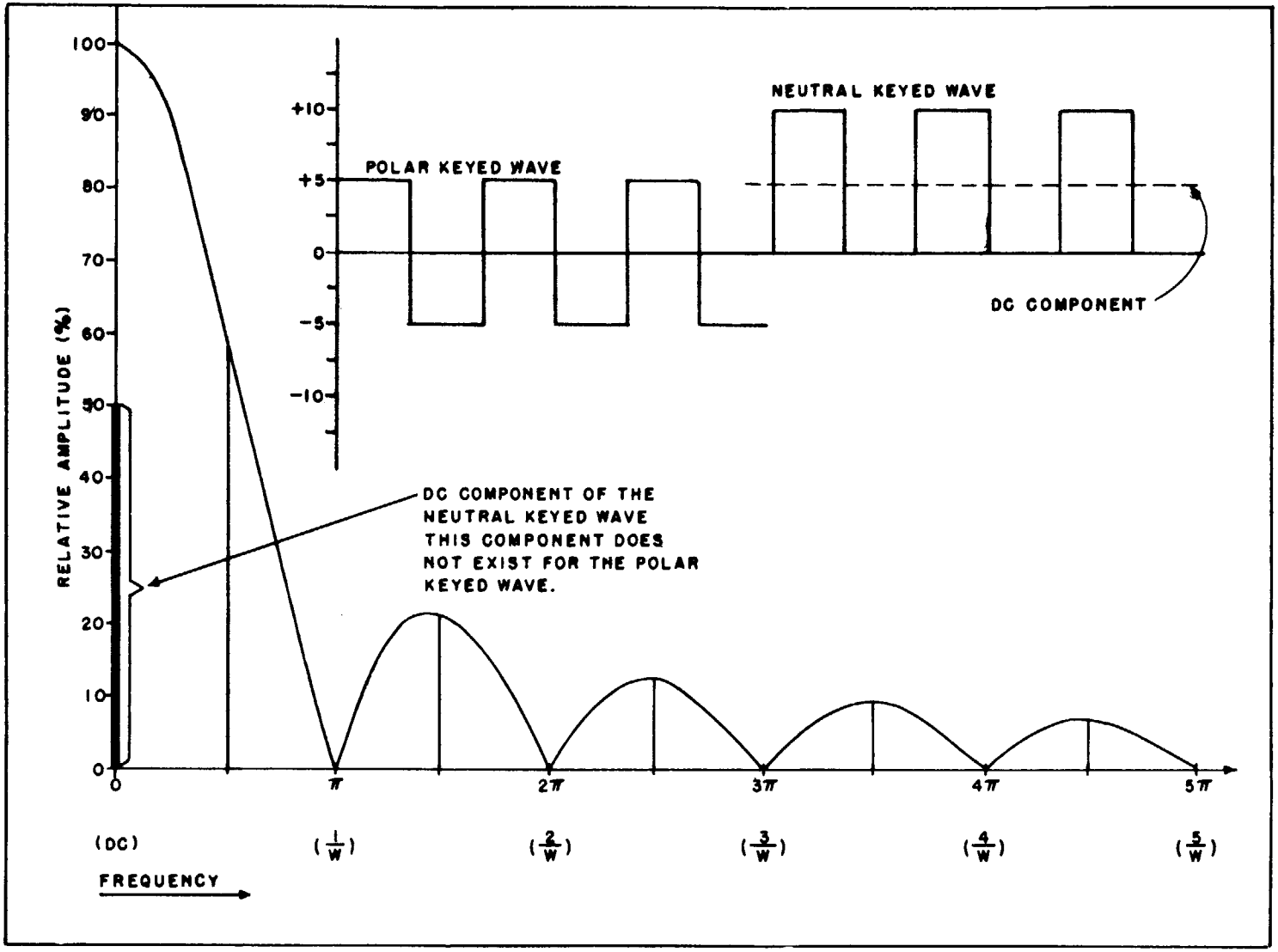


FIGURE 38. EFFECT OF A DC COMPONENT.

DOCID: 3726612

CONCLUSION

By now the student should have become acutely aware that pulses are complex waves and as such their energy is distributed over a band of many frequencies. The student should also be aware that although it is not necessary for equipment to accommodate the entire spectrum of idealized pulses, there is, for any given pulse type, some minimum bandwidth which must be accommodated if the pulse is to serve any useful purpose. In order to place proper emphasis on this bandwidth requirement it was necessary to introduce several new definitions, terms, measurements, and concepts, not the least of which is the function (sine X)/X. Because the rectangular pulse is easy to generate, is frequently encountered, and is convenient to explain, the topics that have been discussed have been centered around this particular type of wave.

There are other things to be remembered also. The highlights among these are spectral lines vs pulse period, dropouts vs pulse duration, the (sine X)/X curve, and transients.

The harmonics of any periodic pulse are multiples of the reciprocal of the pulse period, which turns out to be multiples of the PRF such as 1 x PRF = fundamental, 2 x PRF = second harmonic, 3 x PRF = third harmonic, etc. These harmonics may be represented on a spectrogram by spectral lines, each spectral line indicating by its position and length the frequency and peak amplitude of a sinewave.

Harmonics whose amplitude is zero are those frequencies that correspond to multiples of the reciprocal of the pulse duration, that is multiples of 1/W. These frequencies, by virtue of the fact that their amplitude is zero, are said to "drop out" of the rectangular wave spectrum.

A continuous smooth curve drawn through the peak amplitudes of all the harmonics (including those that drop out) in a rectangular wave spectrogram will result in an envelope that varies in amplitude according to the function (sine X)/X. The importance of any harmonic will depend upon its relative amplitude, and it follows that the harmonics with the greatest amplitudes are contained under the larger lobes of the (sine X)/X curve. There is a little more involved than just passing the larger portion of the energy of the spectrum, that is if wave shape is important. However, the harmonics that will have the predominant effect on the waveshape are those having the larger amplitudes. For a squarewave this usually includes at least the third harmonic with the actual cut off frequency falling at the second dropout (fourth harmonic).

Of the examples given in this lesson one wave stands in a class by itself--the transient. It has no dc level, no PRF, no harmonics, yet its energy has a continuous distribution over the spectrum in such a manner as to follow the same energy

distribution envelope as does a periodic pulse that has the same shape and pulse duration.

All of the foregoing point out that a pulse, whether periodic or transient has its energy distributed over a wide band of the frequency spectrum. Fidelity of reproduction depends upon the amount of this pulse spectrum bandwidth that is passed by the equipment utilizing the pulse. Yet there is more to be said of bandwidth, even the bandwidth of pulses, for nothing has been said of what happens when pulses are used to modulate a carrier. Inasmuch as modulation has not yet been discussed, the bandwidth requirements for pulse modulation of a carrier has been deferred until that time.

IV. PRACTICAL EXERCISES

1. Pulse rise time is measured between

   (a) 0 and 10% of peak amplitude
   (b) 0 and 90% of peak amplitude
   (c) 0 and 100% of peak amplitude
   (d) 10% and 90% of peak amplitude

2. The only wave that always has a 50% duty cycle is a

   (a) rectangular wave
   (b) squarewave
   (c) cosine squared wave
   (d) clipped sawtooth wave

3. Pulse decay time is measured between

   (a) 50% and 0% of peak amplitude
   (b) 50% and 10% of peak amplitude
   (c) 90% and 10% of peak amplitude
   (d) 100% and 0% of peak amplitude

4. The Pulse that best describes the effect of a switch that is turned ON and OFF is a

   (a) rectangular pulse
   (b) stepped pulse
   (c) sinewave
   (d) double pulse

5. The width of a pulse is a measure of its

   (a) time duration
   (b) period
   (c) fundamental and odd harmonics
   (d) rise time of the leading edge

6. A knowledge of the PRF will always indicate the

   (a) pulse duration
   (b) decay time
   (c) duty cycle
   (d) fundamental frequency

7. A certain periodic rectangular wave consists of .5 millisecond elements during which power is ON and 2.5 millisecond elements during which power is OFF. For bandwidth considerations, which element is the pulse?

   (a) it can't be determined from this data
   (b) the 2.5 millisecond element
   (c) the .5 millisecond
   (d) it doesn't matter

8. If conditions were changed in the above example such that the short elements were OFF and the long elements were ON, what then?

   (a) the ON element always determines bandwidth
   (b) it can't be determined from this data
   (c) it doesn't matter
   (d) the shortest element determines bandwidth regardless of which is ON or OFF

9. The idealized perfect squarewave must have

   (a) all harmonics to infinity
   (b) a (sine X)/X dropout at each even harmonic
   (c) all even harmonics to infinity
   (d) a finite rise and decay time

10. Increasing the pulse width

   (a) decreases the required bandwidth
   (b) increases the required bandwidth
   (c) changes the pulse period
   (d) changes the PRF

11. If you had a source of rectangular pulses and you wished to observe the (sine X)/X relationship of the spectral lines the proper equipment to use is

   (a) a spectroscope
   (b) a spectrum analyzer
   (c) a sonograph
   (d) an oscilloscope

12. In the final analysis, what really determines minimum allowable equipment bandwidth?

Questions 13, 14, 15, 16, and 17 are based on the following:

The pulses of a certain pulse train have a pulse duration of 4 microseconds and a pulse period of 10 microseconds.

13. The duty cycle is _____ %

14. The PRF is _____ kilocycles

15. The frequencies of the dropouts are _____

$D.O._1$ = _____

$D.O._2$ = _____

$D.O._3$ = _____

$D.O._4$ = _____

$D.O._5$ = _____

16. On the (sine X)/X curve supplied, draw in all the spectral lines whose amplitude is greater than zero.

17. (a) Which harmonics drop out?

_____ th and _____ th

(b) Do any dropouts occur when there are no harmonics? If so, which ones?

_____ , _____ , and _____ .

DOCID: 3726612

APPENDIX

The (sine X)/X relationship was described in the text as a means of plotting an imaginary envelope which defines the relative amplitude of the spectrum of rectangular waves. Actually the (sine X)/X relationship can be applied to the harmonics of a periodic rectangular pulse without ever using the (sine X)/X curve. This is done by interpreting each of the harmonics in terms of X and then solving the equation $A_N = 2A\ W/P$ (sine X)/X for each. The value of X is $\pi f_n W$, that is π times the frequency of the harmonic times the pulse duration. (The subscript, N, in the equation and n in the expression $\pi f_n W$ simply identifies the harmonic order as A_3 means, "the amplitude of the third harmonic," and f_4 means "the frequency of the fourth harmonic," etc.).

For example, assume a squarewave whose fundamental frequency is 1 kc and whose peak amplitude is 100 volts. Because the wave is square we know that only the odd harmonics have amplitudes greater than zero. These harmonics then are f_1 = 1000 cycles, f_3 = 3000 cycles, f_5 = 5000 cycles, f_7 = 7000 cycles and f_9 = 9000 cycles. The pulse duration will have to be one-half of the pulse period or .5 millisecond.

To solve for the reference amplitude the proper values are simply substituted into the relationship 2A W/P. This gives --

$$2A\ W/P = 2(100)\frac{.5 \times 10^{-3}}{1 \times 10^{-3}} = 100 \text{ volts.}$$

The amplitude of any harmonic, then, is --

$$A_N = 100 \text{ volts (sine X)/X}$$

For the first harmonic (the fundamental), --

$$X = \pi f_1 w = \pi(1 \times 10^{3})\ .5 \times 10^{-3} = \frac{\pi}{2} \text{ radians}$$

When substituted into the function (sine X)/X, the following obtains --

$$\text{(sine X)/X} = \frac{\text{sine } \pi/2}{\pi/2} = \frac{1}{1.57} = .636$$

Then for the fundamental component,

$$A_1 = 2A\ W/P \text{ (sine X)/X} = 100(.636) = 63.6 \text{ volts.}$$

DOCID: 3726612

For the third harmonic

$$X = \pi f_3 w = \pi(3 \times 10^3)\ .5 \times 10^{-3} = \frac{3\pi}{2} \text{ radians}$$

$$(\text{sine } X)/X = \frac{\text{sine } \frac{3\pi}{2}}{3\pi/2} = \frac{1}{9.425} = .212$$

Then for the third harmonic

$$A_3 = 2A\ W/P\ (\text{sine } X)/X = 100(.212) = 21.2 \text{ volts}$$

For the fifth harmonic

$$X = \pi f_5 w = \pi(5 \times 10^3)\ .5 \times 10^{-3} = \frac{5\pi}{2} \text{ radians}$$

$$(\text{sine } X)/X = \frac{\text{sine } \frac{5\pi}{2}}{5\pi/2} = \frac{1}{7.85} = .127$$

Then for the fifth harmonic

$$A_5 = 2A\ W/P\ (\text{sine } X)/X = 100(.127) = 12.7 \text{ volts}$$

For the seventh harmonic

$$X = \pi f_7 w = \pi(7 \times 10^3)\ .5 \times 10^{-3} = \frac{7\pi}{2} \text{ radians}$$

$$(\text{sine } X)/X = \frac{\text{sine } \frac{7\pi}{2}}{7\pi/2} = \frac{1}{10.99} = .091$$

Then for the seventh harmonic

$$A_7 = 2A\ W/P\ (\text{sine } X)/X = 100(.091) = 9.1 \text{ volts}$$

Similarly for the ninth harmonic

$$X = \pi f_9 w = \pi(9 \times 10^3)\ .5 \times 10^{-3} = \frac{9\pi}{2} \text{ radians}$$

$$(\text{sine } X)/X = \frac{\text{sine } 9\pi/2}{9\pi/2} = \frac{1}{14.13} = .0707$$

Then for the ninth harmonic

$$A_9 = 2A\frac{W}{P}\ (\text{sine } X)/X = 100(.0707) = 7.07 \text{ volts}$$

DOCID: 3726612

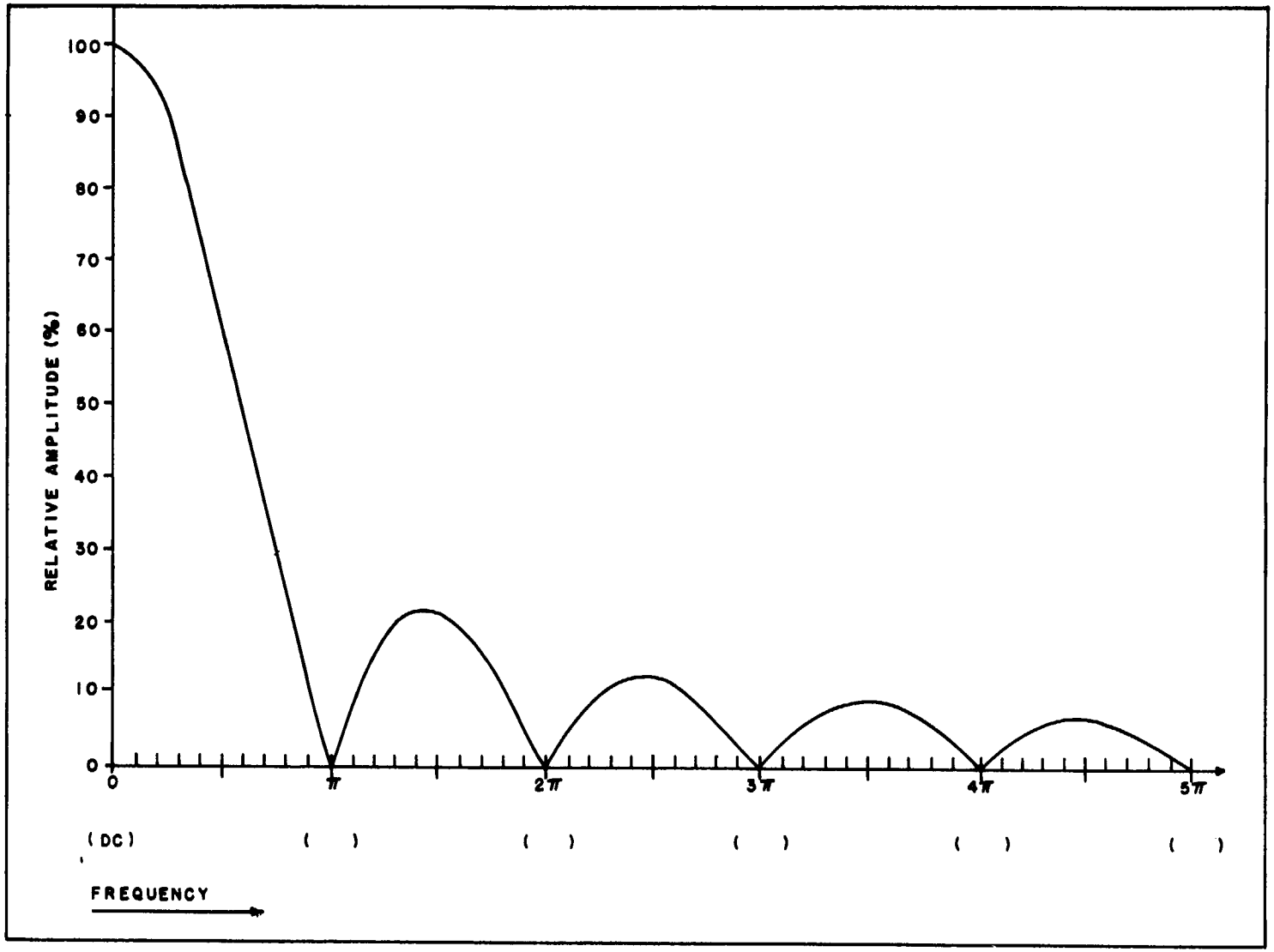


FIGURE 39.

DOCID: 3726612

PART II

MODULATION

THE PROBLEM STATED: The intelligence in an information wave is contained in deviations from the steady state. The aspects of information are amplitude and frequency (period), hence the intelligence in an information wave is contained in deviations from the steady state in either or both of these aspects. Steady, periodic repetition lacks the necessary deviation from the steady state, hence such a wave cannot convey any information either for equipment control or for human interpretation. An example of this is the periodic radio frequency wave. Although it can radiate well, is capable of being propagated over great distances, and can transfer electrical energy between remote points, its usefulness is limited in communication because it cannot contain any information.

Information may be of three types, two of which are distinct classes and the other a combination of the first two. These are discrete information where both message and signal are separate, individual and distinct symbols such as bits, dots, dashes, samples and characters; continuous information, where both the message and the signal are continuous functions of time that can assume any instantaneous value between its maximum and minimum values such as the waves of speech; and mixed information, which consists of both discrete and continuous types.

As it was stated at the beginning, information is aperiodic. By virtue of this, an information wave, whether containing discrete, continuous, or mixed information must also be aperiodic. Although information waves have the aperiodic deviations in amplitude and frequency that are necessary to carry intelligence, these waves (as they are produced by their sources) generally fall in a part of the frequency spectrum that is below the radio frequency range and are incapable of radiation by any practical means.

Of whatever type the information wave may be, it must come from some information source. An information source is simply a device that produces messages. Persons and telemetry end-instruments, are typical information sources. From any source, information is rather useless unless it can be transmitted by some means from the source to the destination. (This is the problem suggested by an R-F wave that can radiate but cannot carry intelligence and an information wave, that can carry intelligence but cannot radiate.)

BRIDGING THE GAP: To make the passage from source to destination necessitates that the information travel or be propagated through some medium. The medium through which, or by which it travels may be wire, beam of light, radio frequency wave, or some other. Since the information is not always directly applicable to the medium necessary or available for its transmission, it becomes necessary to

DOCID: 3726612

encode the information from the information source into a form of message-bearing signal that can be transmitted through the required medium.

THE PROBLEM SOLVED: Encoding information in the manner, and for the purpose, just indicated is called modulating or the process of modulation. This encoding or modulating process is all important because it overcomes the difficulties (such as an R-F wave that can radiate but cannot carry intelligence, and an information wave that carries intelligence but cannot radiate) that would otherwise make our modern methods of communication impossible and it enables information to travel from source to destination without intervening circuitry apart from that which is in the actual transmitter and receiver. This is accomplished when the radiating ability of the periodic R-F wave and the characteristic aperiodicity of the information wave are combined. Modulation does this by taking the energy that can be radiated through space and producing variations in it that correspond to the variations in the information wave. The resultant complex wave that is produced is otherwise periodic except for some aspect which is varied in synchronism with and in proportion to the instantaneous variations in frequency and amplitude of the information wave.

From the modulation process two benefits accrue: first, the information is translated to a higher portion of the spectrum where it can radiate, and secondly, it is no longer limited to the power of the information source, since the translated frequencies derive their power from the encoder (modulator) itself.

RECAPITULATION: The study of modulation was deferred until such time as sufficient background in the basic nature of simple harmonic waves and complex waves and their spectra had been presented. This done, the introduction of the subject of modulation is now appropriate.

The standard definition of modulation is, "The process or result of the process whereby some characteristic of one wave is varied in accordance with another wave" (53 IRE 11-S1). Basically, the term, modulate, means regulate. Considering the basic meaning as applied to the standard definition, it evolves that certain characteristics of one wave (called the modulated wave) are regulated in some manner that directly depends upon the instantaneous amplitude of another wave called the modulating wave.

In an earlier assignment, the sinewave was introduced and various sorts of complex waveforms were developed from these sinewaves. This was done by simply adding the instantaneous amplitudes of concurrent sinewaves. Preparatory to developing complex waves from sinewaves, considerable time was spent on the sinewave itself: its period, frequency, amplitude, phase and true shape. Here the sinewave formula was introduced and a sinusoidal cycle was plotted by use of that formula. Reviewing briefly, the student may recall the formula as, $a = A \sin \theta$,

DOCID: 3726612

where a represents any instantaneous amplitude of the resultant sinewave, A is the peak amplitude of the wave (or length of the rotating vector that generates it), and θ is the angle made by the vector at the particular instant of time. It should be remembered that the length, A, of the vector when multiplied by the sine of the instantaneous angle, θ, resulted in the instantaneous amplitude, a.

KINDS OF MODULATION: In the study of modulation, the sinewave formula is again an important item. In it are to be found the characteristics of a sinewave that are subject to modulation. It will be noted that there are only two such characteristics--the amplitude, A, and the angle, θ. The sinewave formula, therefore, limits the kinds of modulation to two: amplitude modulation and angle modulation.

PLAN OF DISCOURSE: The two forms might ordinarily constitute the subject matter for two corresponding discussions on modulation but there is a modified form of amplitude modulation that warrants separate treatment: it is single sideband. Therefore three topics--amplitude modulation, single sideband and angle modulation--will occupy the discussion in this and the following two volumes of this study guide series.

DOCID: 3726612

CHAPTER 3

AMPLITUDE MODULATION

Of the two basic types of modulation, amplitude modulation, abbreviated A-M, is most familiar to the public in general because it is the system used in standard broadcasting, and indeed has been since A-M broadcasting began as a regularly scheduled public service in 1920. It is not only the oldest form of modulation historically but it is the form of modulation most easily produced, and that form which is most frequently encountered in nature. As its name implies, in A-M, amplitude is the wave characteristic that is encoded to carry the frequency and amplitude aspects of an information wave.

It is our purpose in this chapter to define amplitude modulation, establish the possibility of its existence, observe its waveform, examine its spectral structure and required bandwidth, and to present the general terminology used to describe the A-M phenomenon.

I. TERMS AND DEFINITIONS

The terms and definitions that are employed in amplitude modulation are few and simple. They are presented early in this chapter to give the student ample time to familiarize himself with them.

A. Modulating Wave

A wave whose instantaneous values are encoded by producing variations that are in proportion to and in synchronism with it in some other wave, is called a modulating wave. A modulating wave may or may not be sinusoidal and it may be periodic or aperiodic. Whatever its type or purpose it becomes a modulating wave by virtue of the manner in which it is used in a modulator, namely to produce modulation. Theoretically, a modulating wave may exist in any part of the frequency spectrum. The principle reasons for using a wave as a modulating wave are to permit multiplexing (covered in a later lesson) and/or to permit the radiation and propagation of its variations through space. Figure 40A illustrates a sinusoidal modulating wave. (In this lesson the peak amplitude of a sinusoidal modulating wave will be abbreviated, A_m).

B. Carrier Wave

Another wave used in the modulation process is used to encode the modulating wave. In general, it too may exist in any part of the frequency spectrum so long as its variations are substantially higher in frequency than those of the modulating wave. This wave, although it doesn't actually carry anything either before or after modulation, is called the carrier wave. Only sinusoidal carriers

DOCID: 3726612

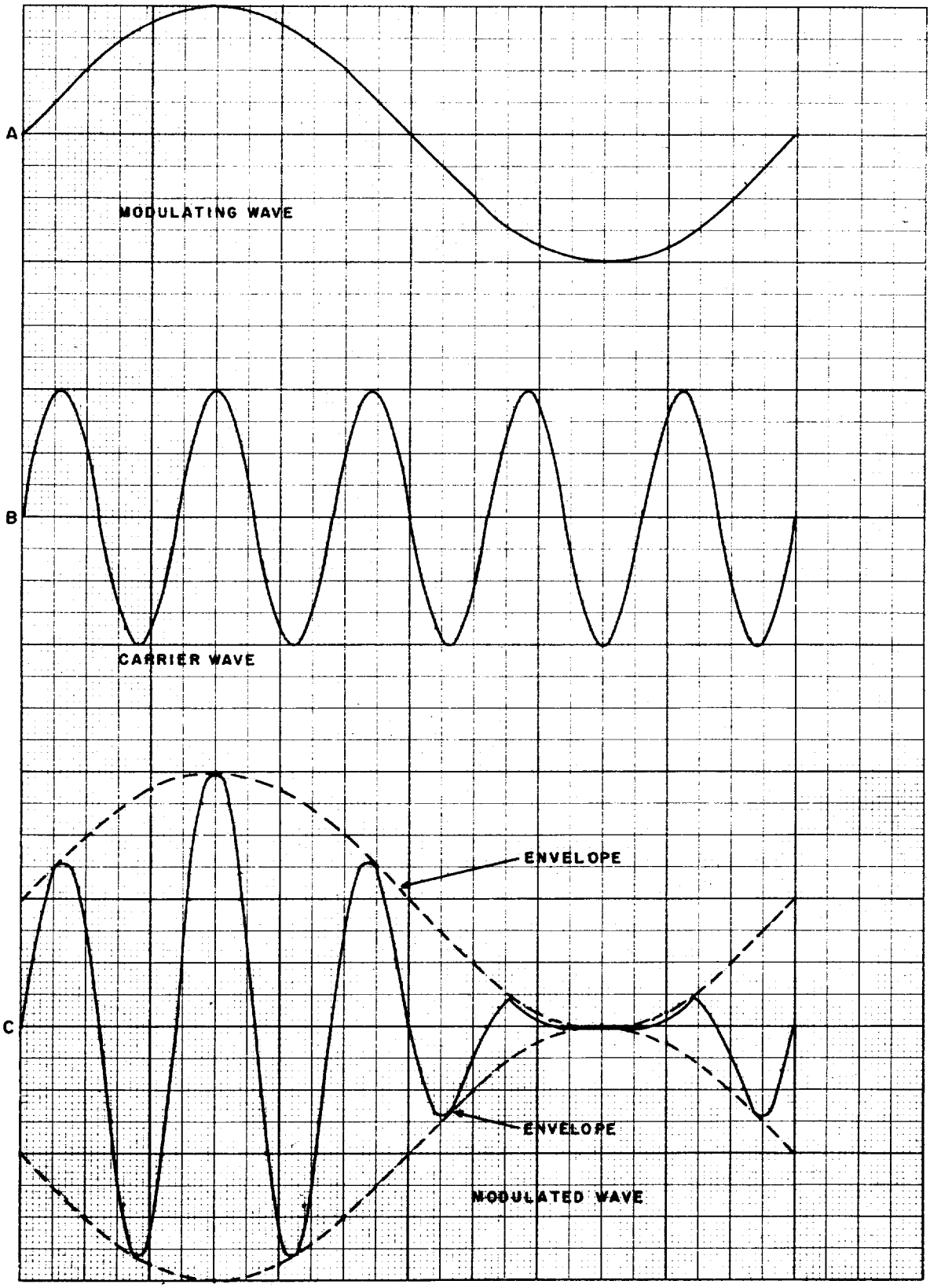


FIGURE 40. RELATIONSHIP OF CARRIER WAVE. MODULATING AND MODULATED WAVE.

DOCID: 3726612

will be considered here, however, one form of multiplexing (to be discussed in a later lesson) makes use of pulse carriers. Figure 40B illustrates a carrier wave. (In this lesson, the peak amplitude of a sinusoidal carrier wave will be abbreviated, A_c).

C. Modulated Wave

When the variations in a modulating wave are encoded in a modulator by the carrier wave, the resultant is called a modulated wave. The modulated wave is always a complex wave that contains variations that are in synchronism with, and in proportion to variations in the modulating wave. Figure 40C illustrates a modulated wave resulting from amplitude modulation.

D. Amplitude Modulation

Figure 40B presents a sinusoidal modulating wave, a carrier, and the complex wave resulting from amplitude modulation. The appearance of the complex wave is described well by a widely used non-mathematical definition which states simply that amplitude modulation is, "Modulation in which the amplitude of a carrier is the characteristic that is varied" (53 IRE 11-S1). This definition is simple and concise and when it is compared with Figure 40, its correctness is confirmed for every detail of the appearance of the modulated wave. In fact the modulating wave and its mirror image defines an imaginary boundary called an envelope (indicated by dashed lines in Figure 40C) that appears to determine the amplitude of the carrier positively and negatively. From the standpoint of appearance this definition is good, but from the standpoint of reality it is incorrect and is at variance with the definitions previously given which speak of a complex modulated wave resulting from the modulation process, not of a carrier whose amplitude is varied.

1. Language of Appearance

Probably every reader of this study guide has spoken of the sun rising and of the sun setting. Equally, it is doubtful if anyone of the readers believes that the sun rises or sets. We all realize that it is actually the rotation of the earth on its own axis, and not an actual rising and setting of the sun that accounts for the observed phenomena in early morning and in late evening. Yet, the expression sunrise and sunset are used by the scientific and the unscientific alike because it is handy to use and it describes what appears to be happening.

2. Justification of the Definition

The definition just given for amplitude belongs to the same

DOCID: 3726612

category as the expressions "sunrise" and "sunset"--it is the language of appearance! As such, it describes at a glance how the modulated A-M waveform looks but it does not tell what actually happens. Unfortunately, most popular definitions of amplitude modulation share in this same shortcoming; however, the language of appearance need not be harmful if it is realized by its users that there is more to modulation than meets the eye.

3. Mathematical Approach

By mathematics, it is possible to define amplitude modulation accurately and precisely but no attempt will be made to do so here (equations that form the basis of such a definition appear in the appendix). Instead, the results of such a definition will be realized when an A-M modulated waveform is analyzed into a spectrum consisting of carrier and sidebands with appropriate amplitude and phase relationships.

E. Modulation Factor

A modulating wave that has a low amplitude will produce a smaller amplitude variation in the modulated wave than a high amplitude modulating wave will. This fact gives rise to a need for expressing the degree (or depth) of modulation produced by a wave of some particular amplitude. This is expressed by a ratio called the modulation factor, M_a. The modulation factor is simply a ratio of the peak amplitude variation used (A_m) to the maximum design variation (A_c). Under proper operating conditions A_m will always equal or be less than A_c, therefore, the modulation factor, $M_a = \frac{A_m}{A_c}$, will not be allowed to exceed unity. This factor varies from 0 to 1 decimally and expresses the fractional change in amplitude of the modulated wave due to modulation. Figure 41, illustrates a 100% modulated wave and indicates where the values of A_m and A_c are measured.

II. MATHEMATICAL POSSIBILITY OF A-M MODULATION

The entire realm of possible modulations of a sinewave carrier is wrapped up in the brief sinewave formula that the student learned earlier in this course. That formula is, $a = A \sin \theta$. The student should already be familiar with the significance of this formula since practice has been given in its use.

A. The Interpretation of θ

For purposes of understanding modulation it is appropriate to subject θ to a little closer examination. When the student plotted sinewaves from

DOCID: 3726612

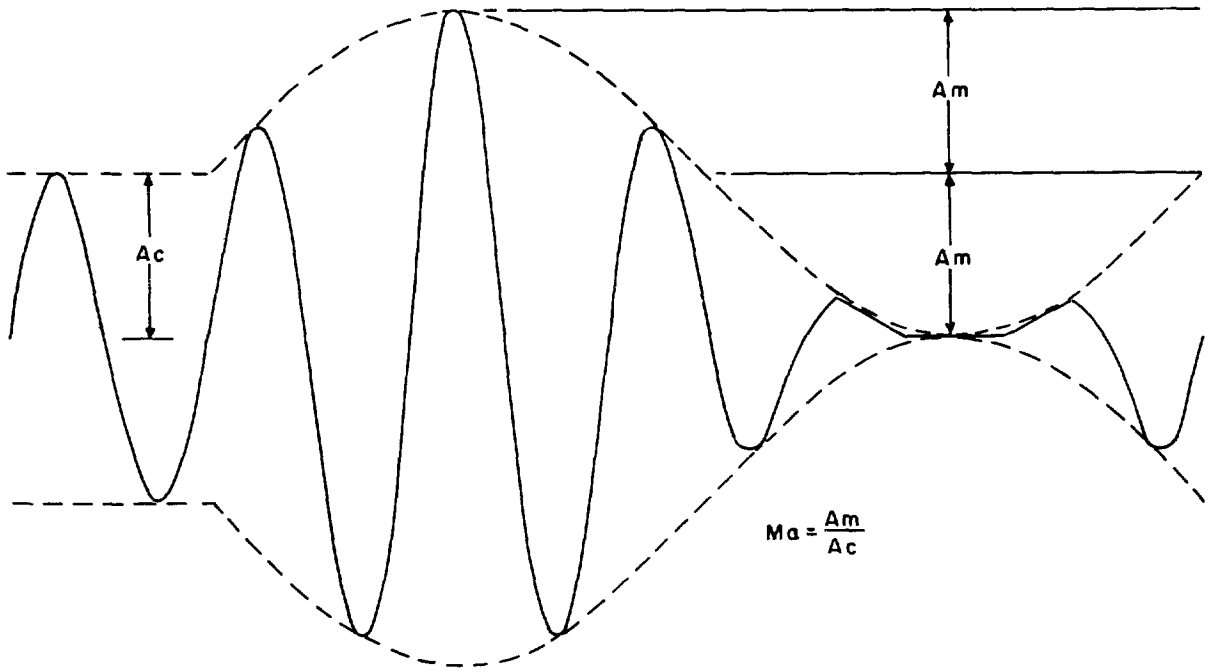


FIGURE 41. HOW THE MODULATION FACTOR IS DETERMINED.

the formula, instantaneous values for θ were arbitrarily selected as 0°, 30°, 60°, 90°, etc. . It was later found that angles expressed in degrees could also be expressed in radians by use of the degree-to-radian conversion factor, $2\pi/360^{o}$. Therefore, if desirable, θ could be expressed in radians as easily as in degrees.

A more sophisticated representation of θ is $(2\pi ft + \phi)$ which is an expression in radians. Using this interpretation of θ, the sinewave formula becomes,

$$a = A \sin (2\pi ft + \phi)$$

(where ϕ is the Greek letter phi). That θ could represent a long expression such as $2\pi ft + \phi$, or that such an expression could even be an angle may seem strange at first, but with a brief but simple explanation everything should fall properly in line. If ϕ were equal to zero in a given case, θ would simply equal $2\pi ft$. If we make the assumption that $\phi = 0$ we have found a good place to begin.

1. Angular Frequency

Angular frequency considers only the first three factors in the expression $2\pi ft$, namely the $2\pi f$. It was shown in Waveform

DOCID: 3726612

Analysis that a cycle consists of 2π radians or 360°. This being true it can be said that there are,

$$\frac{2\pi \text{ radians}}{\text{cycle}} \quad \text{or} \quad \frac{360 \text{ degrees}}{\text{cycle}}$$

The first of these two expressions is represented by the "2π" part of the angular frequency, $2\pi f$.

When a cycle is completed in a certain period of time, the rate at which a given number of cycles is completed is defined as frequency and is expressed in cycles/second. The factor "f" in the expression $2\pi f$ simply stands for frequency.

Suppose, now, that 2π (which represents $\frac{2\pi \text{ radians}}{\text{cycle}}$) were multiplied by f (which represents $\frac{\text{cycles}}{\text{second}}$). There would obtain,

$$2\pi \times f = \frac{2\pi \text{ radians}}{\cancel{\text{cycle}}} \times \frac{\cancel{\text{cycles}}}{\text{second}} = \frac{2\pi \text{ radians}}{\text{second}} = 2\pi f$$

This is all that is indicated by the factors $2\pi f$. Multiplying cycles per second by 2π simply converts cyclic frequency to angular frequency, that is, $2\pi f$ tells the rate at which complete cycles of 2π radians are being developed.

Some prefer to think of $2\pi f$ as an angular velocity instead of an angular frequency. This is perfectly allowable because, since $f = \frac{1}{T}$, by substitution $2\pi f = 2\pi \left(\frac{1}{T}\right) = \frac{2\pi}{T}$. This latter form has the qualities of a velocity (like $\frac{\text{miles}}{\text{hours}}$) since it expresses the rate at which distance (2π radians is angular distance) is covered. Whether one prefers to think of $2\pi f$ as angular frequency or angular velocity makes little difference--the result is the same. Very frequently the angular rate, $2\pi f$, is abbreviated by the Greek letter ω (omega).

2. Angular Distance

If you were to drive at a constant velocity of 60 miles/hour for a long period of time the distance covered during equal time intervals within that period would be equal. Thus, after 10 minutes (1/6 of an hour) 10 miles would have been covered, after another 10 minutes another 10 miles would have been covered, and so on.

DOCID: 3726612

In other words as long as the velocity is constant the distance covered increases uniformly with time and a constant velocity yields a linearly increasing distance that is proportional to the elapsed time. For example, for a velocity of 60 miles/hour, the total distance after one hour is given by,

$$\frac{60 \text{ miles}}{\text{hour}} \times 1 \text{ hour} = 60 \text{ miles.}$$

Angular velocity can be treated in the same manner. If the angular distance of a rotating vector is increasing at a certain constant angular velocity measured in radians/second over a long period of time, the angular distance covered during equal time intervals within that period will be equal, i.e., the angle increases by equal amounts for equal intervals of time. As long as the angular velocity is constant the angular distance covered increases uniformly with time and a constant angular velocity yields a linearly increasing angular distance that is proportional to the elapsed time. For example, for an angular velocity of 2π radians/second, the total angular distance after one second is given by,

$$\frac{2\pi \text{ radians}}{\text{second}} \times 1 \text{ second} = 2\pi \text{ radians.}$$

The obvious conclusion, then, is that 2π ft is simply the total angular distance that has been travelled by a rotating vector in the amount of time indicated by t. If ϕ were zero as was assumed earlier, then the angle θ is defined completely by 2π ft. That is, θ expressed in radians is an angle whose size depends upon frequency (f) and elapsed time (t).

One illustration will be offered to crystallize the explanation: Assume a frequency of one kilocycle. What is the value of θ in radians 250 microseconds after the beginning of a cycle? How many degrees is this?
Solution:

$$\theta = 2\pi \text{ ft radians}$$

$$\theta = \frac{2\pi \text{ radians}}{\cancel{\text{cycle}}} \times \frac{1 \times 10^{3} \cancel{\text{ cycles}}}{\cancel{\text{second}}} \times \frac{.25 \times 10^{-3} \cancel{\text{ seconds}}}{1}$$

$$\theta = .5\pi \text{ radians}$$

DOCID: 3726612

Expressed in degrees:

$$\theta = .5 \cancel{\pi} \text{ radians} \times \frac{\overset{180}{\cancel{360}} \text{ degrees}}{\cancel{2\pi} \text{ radians}} = 90 \text{ degrees.}$$

3. The Effect of an Added Phase Angle

In order to clear away obstacles in explaining the angle θ, it was assumed that the angle ϕ in the expression, $\theta = (2\pi \text{ ft} + \phi)$ was equal to zero so it could be neglected. However ϕ is not always equal to zero. If there is any departure from the uniformly increasing phase angle, 2π ft, this phase difference is represented by ϕ. Depending upon whether ϕ is positive or negative, when added to 2πft it may increase or decrease the magnitude of θ. (And of course, to make a paper addition of ϕ to 2π ft, ϕ must be expressed in radians.)

B. The Full Sinewave Equation

The expanded sinewave equation should seem a little less difficult now than it did a few paragraphs previous since each of the symbols in the equation, $a = A \sin(2\pi \text{ ft} + \phi)$ has been assigned some meaning.

1. Variables and Constants

In this equation certain quantities can be varied such as the peak amplitude (A), the frequency (f) and the phase angle (ϕ). These quantities are called variables. There are other quantities that cannot undergo any variation at any time such as 2 which is always 2, π which is always equal to 3.141592..., or the sine of any fixed angle. These quantities are called constants. Although time varies (increases) continuously any fixed interval of time such as one second or one minute has a constant magnitude and as such is a constant.

2. The Mechanism of Modulation

The variables, A, f, and ϕ in the sinewave equation make it vulnerable to distortion. If any attempt is made to vary or suppress any of these quantities distortion is the immediate and inevitable result. If there is a controlled distortion that varies in accordance with the variations in a modulating wave, the distortion is called modulation, the resultant distorted wave is called a modulated wave

DOCID: 3726612

and the sinewave whose characteristics are acted against is the carrier.

It should be emphasized at this point that only a sinewave has a single, constant frequency--but distortion indicates a complex wave which is something more than a single constant frequency. So, in the effort to suppress any of the variables in a sinewave carrier in accordance with a modulating wave, there is no longer the pure frequency of the carrier but a mixture of the original and new frequencies. The particular new frequencies that are developed will depend upon both the type of modulation and the method employed in producing it. The useful new frequencies, (which are the only ones of interest here) are called side-frequencies or sidebands and will be discussed shortly.

3. Modulation Possibilities

The tabulation below illustrates the type of modulation that results from the attempt to vary the characteristics of a sinewave carrier.

$a = A \sin (2\pi ft + \phi)$
Amplitude Modulation

$a = A \sin (2\pi ft + \phi)$
Frequency Modulation

$a = A \sin (2\pi ft + \phi)$
Phase Modulation

Note that any change in f or ϕ is simply a variation in the angle θ since $\theta = 2\pi ft + \phi$. Therefore, frequency modulation and phase modulation are both classified as angle modulation.

III. AMPLITUDE MODULATION WAVE FORM

It has been shown previously that when sinewaves of various frequencies, phases and amplitudes are combined they produce a complex wave. This mixing may be either of two types--linear or non-linear.

Linear mixing is called adding. Figure 42, illustrates two sinewaves (the same two shown in Figure 40 as a carrier and a modulating wave) when they are simply added or linearly mixed. Since the resultant complex wave in Figure 42

DOCID: 3726612

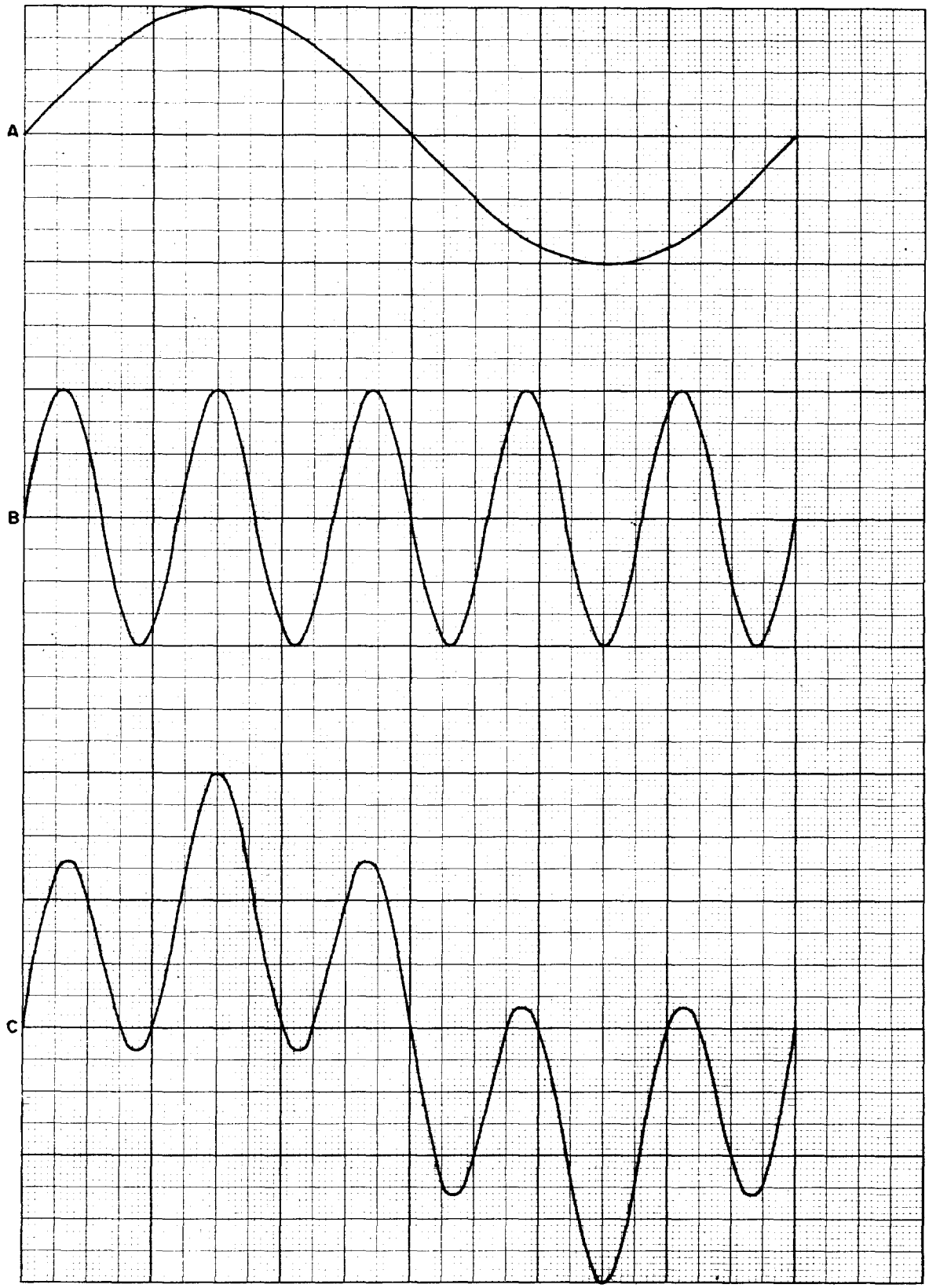


FIGURE 42. TWO SINEWAVES OF UNLIKE FREQUENCY LINEARLY MIXED,

DOCID: 3726612

differs greatly from the resultant complex wave in Fig. 40, then obviously amplitude modulation and linear mixing are not the same.

Non-linear mixing is most frequently called heterodyning. Amplitude modulation is a result of heterodyne action (the production of sum and difference frequencies from two or more parent frequencies). The whole difference between the resultant complex waveform in Fig. 40 and that in Fig. 42 is the type of mixing. The latter is linear, the former is non-linear. It is the simple addition of the instantaneous amplitudes of the original components in contrast to heterodyne action whereby beat frequency components (sum and difference) are created.

A. Generation of the A-M Wave

1. Heterodyne Action

Heterodyne action is the effect produced by two or more frequencies interacting to produce new frequencies called beat frequencies that are equal to the sum of the original frequencies and the difference of the original frequencies.

When two or more frequencies are linearly mixed (i.e., added) a varying amplitude may be produced at a beat frequency rate (a single difference in this case) but no beat frequency components (sum or difference frequencies) are produced because the simple addition of waves does not produce any heterodyne action. The production of beat frequency components requires that the original waves (modulating wave and carrier) be distorted as well as added.

2. Distortion

Distortion, it was explained in an earlier treatment, is due to circuit non-linearities. It was found that when a pure sinewave was distorted harmonics of that sinewave frequency were created. The same non-linearities that will distort a single pure sinewave to produce new frequencies that are integral multiples (harmonics) of the original frequency, will, when acting on two or more sinewaves simultaneously, distort them to produce additional new frequencies corresponding to the sums and differences of the fundamentals and harmonics of the applied sinewaves.

3. Modulation

Practical modulators are designed to suppress the generation of harmonics since such energy represents unwanted distortion and

DOCID: 3726612

is in no way essential to the desired modulation. With the undesired distortion products suppressed, the only energy extracted from the modulator is that of the carrier and, for each modulating frequency, a pair of side frequencies (the beat frequencies) which straddle the carrier frequency. The upper side frequency is equal to the sum of the modulating sinewave frequency and the carrier frequency, and the lower side frequency is equal to the difference of the modulating frequency and the carrier frequency. The carrier and the two side frequencies are illustrated in Fig. 43 as the exclusive components of the A-M complex wave.

4. Side Frequencies

The sum and difference frequencies when not referred to a carrier are called beat frequencies, but when they are referred to the carrier they are called side frequencies because they occur at the side of the carrier when the modulation spectrum is observed.

B. Analysis of the A-M Wave

The existence of side frequencies and the absence of the modulating wave in the complex modulated wave are physical realities, the evidence of which can easily be established by use of appropriate filters and amplifiers. Also it is possible to establish physically that the carrier itself is ultimately unaffected by the process of amplitude modulation. Mathematics, of course, can show these facts, but since such mathematical proof is not within the scope of this course it has been deferred to the appendix for the benefit of any who care to pursue the subject to that extent. Without the use of either mathematics or equipment, however, the student should be able to prove to his own satisfaction that the modulation waveform is the sum of the instantaneous amplitudes of the carrier and side frequencies as illustrated in Figure 43.

1. Side Frequencies and Sidebands

So far the term side frequencies has been used when, perhaps, some readers who have had some previous familiarity with the subject are better acquainted with the term, sidebands. The choice of terms used here, however, has been deliberate and with reason. Certain facts which were stated earlier in the course will be repeated and should establish firmly the reason behind the present choice of terms.

Band, as applied to frequency, implies a range of frequencies.

DOCID: 3726612

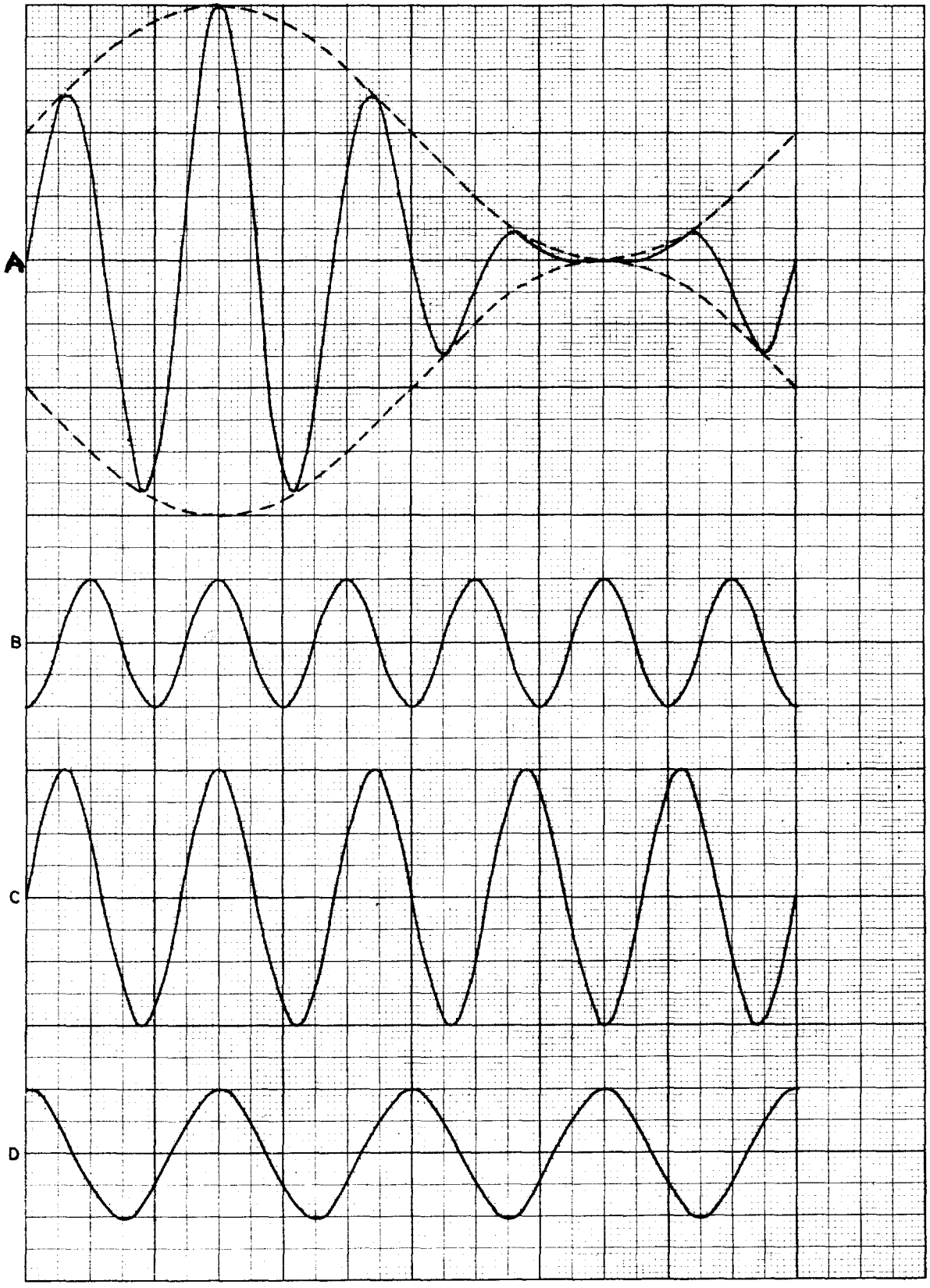


FIGURE 43. A.M. WAVE COMPONENTS.

DOCID: 3726612

For this reason a single frequency cannot be thought of as a band. The extent of any band of frequencies is bounded by the lowest and highest frequencies of the band. The frequency difference between the lower and upper limits of a band is defined as the bandwidth. It should also be recalled that a single frequency cannot have bandwidth--a fact easily understandable because the upper and lower limits of a single frequency are one and the same, hence there is no width between the limits.

It is in the light of these facts that when a single, discrete sum or difference frequency is considered, the term side frequency is appropriate rather than side band. It is allowable, though to refer to the region bounded by the carrier and the lowermost or uppermost side frequency as a sideband, whether the bands are occupied by any other frequencies or not. Any existing side frequencies (but not the carrier) are properly termed sideband components.

2. Power in an A-M Wave

There is more power in an A-M wave than there is in an unmodulated carrier. The power in the A-M modulated wave consists of power supplied by the carrier plus the total power contained in the side frequencies. The carrier power after modulation is the same as before modulation, that is, it is unchanged. The power in the sideband components is contributed by the modulating wave. The total power in the A-M modulated waveform is related to the carrier power by the modulation factor as shown in the equation (Figure 45).

$$P_m = P_c \left(1+\frac{M_a^2}{2}\right)$$

where P_m is the power in the modulated wave and P_c is the power in the carrier. For example, when the carrier power is 200 watts and the degree of modulation is 100% (i.e., m = 1) the power in the modulated wave is

$$P_m = 200 \left(1+\frac{1^2}{2}\right)$$

$$= 200 \left(1+\frac{1}{2}\right) = 200 \ (1.5)$$

$$= 300 \text{ watts.}$$

DOCID: 3726612

(NOTE: In the succeeding pages a number of spectra appear. These spectra represent relative voltage or current and will not satisfy the power equation just given. If the values on these spectra are squared they will then satisfy the power relationship.)

Notice that for 100% modulation one-third of the total power in the modulated wave is contributed by the side frequencies, or stated in other words, where the modulation is 100% the side frequencies must supply power equal to one half the carrier power. When the degree of modulation is 100%, the complete relationship between carrier power, power in the modulated wave, and total sideband power can be summarized by

$$P_{SB} = \frac{1}{2}P_c = \frac{1}{3}P_m$$

It should not be overlooked that one-half of the total sideband power is contributed by the components of the lower sideband, and the other half is contributed by the components of the upper sideband.

3. Frequency Translation

It has already been suggested several times that there is no energy in the A-M modulated waveform at the frequency of the modulating wave although the envelope varies with the amplitude of the modulating wave. The energy originally in the modulating wave is represented by energy at the sideband components hence, the energy in the side frequencies is proportional to the energy that was in the modulating wave. Because of this relationship between the modulating wave and the side frequencies, any variation that is made in the modulating wave will occur in the side frequencies. For example, if the amplitude or the frequency of the modulating wave were to be increased or decreased, corresponding changes would occur in the side frequencies. The ultimate result is that by means of modulation it is possible to translate variations in a low frequency modulating wave into corresponding variations in a high frequency wave (the side frequencies).

IV. AMPLITUDE MODULATION SPECTRUM

It has been shown that when modulating with a single frequency, a modulated wave--complex in composition--is generated. This complex wave consists of three frequencies: the carrier, a sum frequency (upper side frequency)

DOCID: 3726612

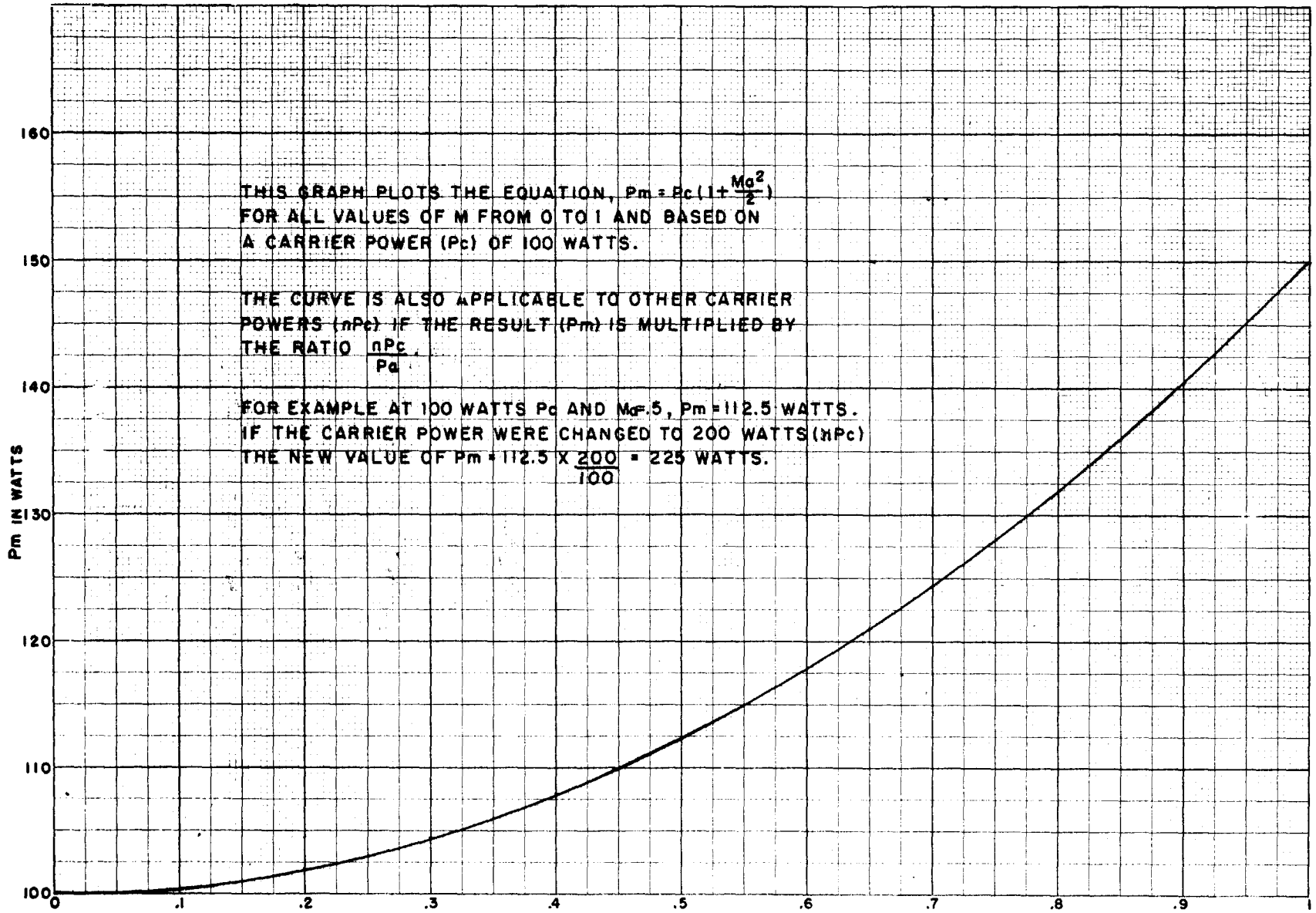

THIS GRAPH PLOTS THE EQUATION, $Pm = Pc(1+\frac{Ma^2}{2})$ FOR ALL VALUES OF M FROM 0 TO 1 AND BASED ON A CARRIER POWER (Pc) OF 100 WATTS.
THE CURVE IS ALSO APPLICABLE TO OTHER CARRIER POWERS (nPc) IF THE RESULT (Pm) IS MULTIPLIED BY THE RATIO $\frac{nPc}{Pc}$.
FOR EXAMPLE AT 100 WATTS Pc AND Ma=.5, Pm=112.5 WATTS. IF THE CARRIER POWER WERE CHANGED TO 200 WATTS (nPc) THE NEW VALUE OF Pm = 112.5 X $\frac{200}{100}$ = 225 WATTS.
Pm IN WATTS
100
110
120
130
140
150
160
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1

FIGURE 44. A-M MODULATION FACTOR (Ma)

DOCID: 3726612

and a difference frequency (lower side frequency). These three components of the modulated wave constitute the spectrum of the A-M modulated wave when the modulating wave is but a single frequency.

When the modulating wave consists of more than a single frequency (which is the usual case), for example, a voice wave or other complex wave, each of the components of the complex modulating wave must create its own pair of side frequencies during the modulation process. All such components that are generated and the carrier make up the spectrum of the A-M modulated wave when the modulating wave comprises more than one frequency.

A comparison of the spectrum presentations when both complex waves and single sinusoids make up the modulating wave will, perhaps, clarify the relationship of carrier and sideband components.

A. Spectrum for Single Tone Modulation

The relationship between the carrier and side frequencies generated when modulating with a single sinewave was illustrated in Figure 43. That these components of the resultant modulated wave can be resolved into a spectrum is shown in Figure 45. Notice that each sinusoidal component of the modulated wave of Figure 43 is represented by a spectral line in Figure 45.

If it is assumed that the carrier in Fig. 43, 45 and 46 is 5 kilocycles per second, and that the modulating wave is 1 kilocycle per second, the lower side frequency will be 5KC-1KC = 4KC and the upper side frequency will be 5KC +1KC = 6KC. The side frequencies, then, are separated from the carrier by a band equal to the modulating frequency which in this instance is one kilocycle.

It should be noted that the modulating wave is represented twice in the A-M modulated wave: once by the lower side frequency and again by the upper side frequency. It should also be observed that the modulating wave has given one-half of its amplitude to the upper side frequency and one-half to the lower side frequency. The signal bandwidth of the modulated wave extends from the lowest side frequency (4KC in this instance) to the highest side frequency (6KC). Although the presence of both sideband components represents a duplication of information, in ordinary amplitude modulation it is necessary that the circuit bandwidth of an equipment be at least as wide as the signal bandwidth of the modulated wave if the waveform is to be passed without distortion or loss of information.

B. Spectrum for Multitone Modulation

The basic difference between using a sinewave for a modulating

DOCID: 3726612

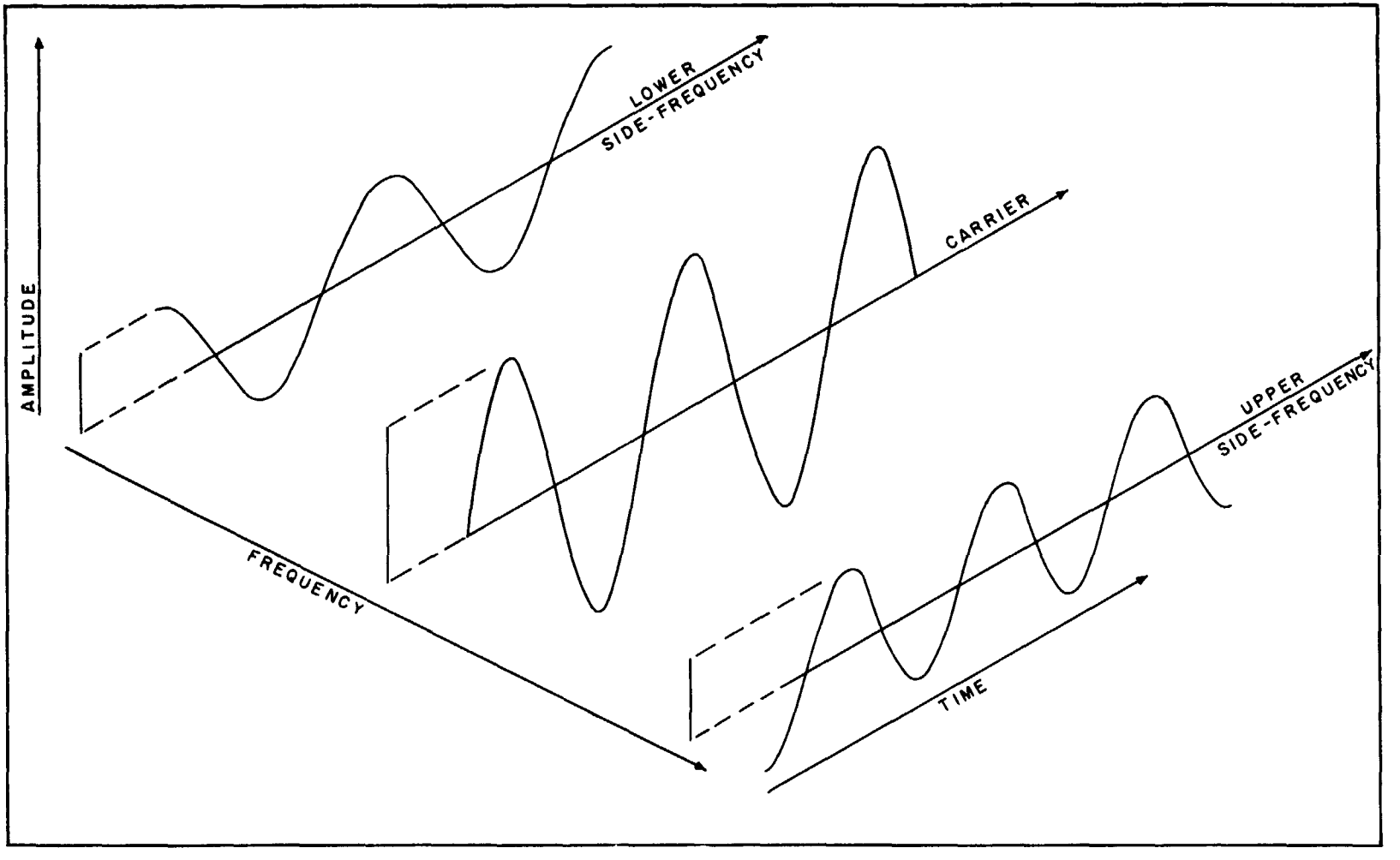


FIGURE 45. COMPARISON OF A-M COMPONENTS AND SPECTRA.

DOCID: 3726612

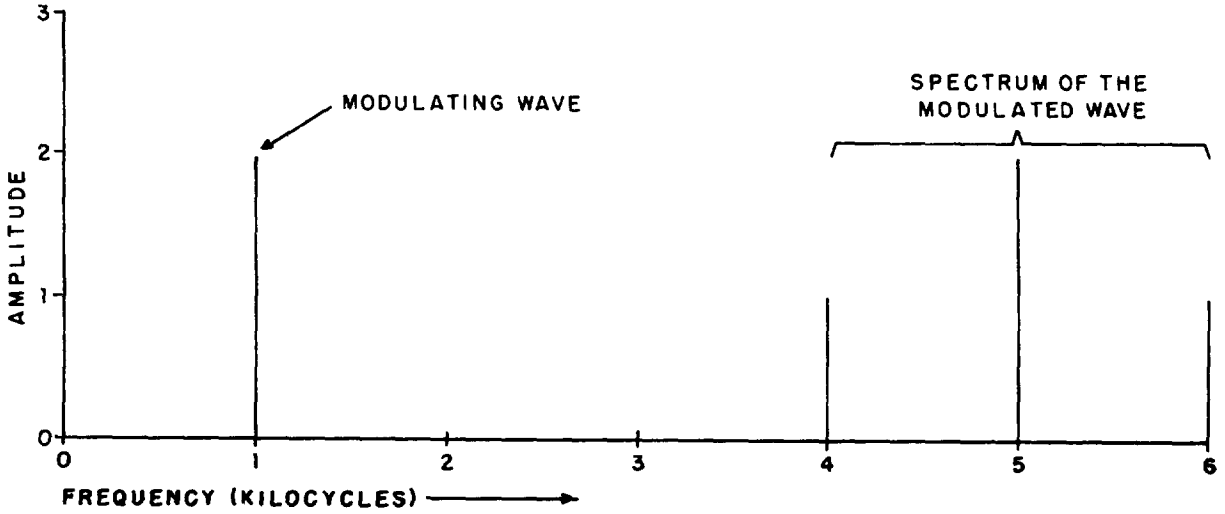


FIGURE 46. COMPARATIVE SPECTRA OF THE MODULATING AND MODULATED WAVES.

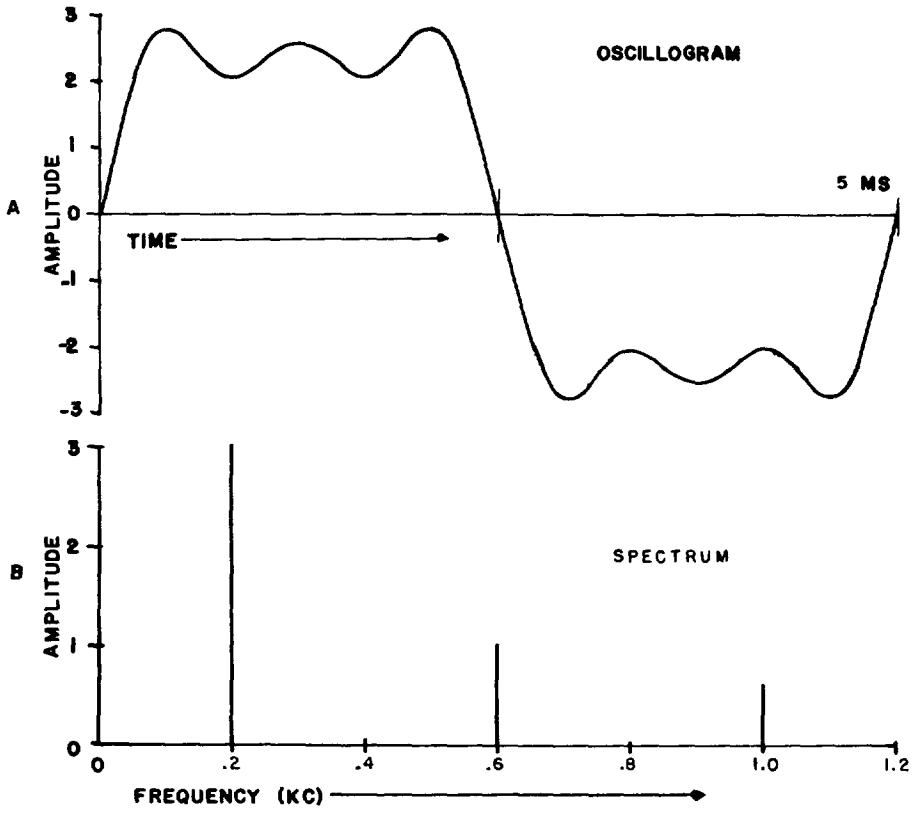


FIGURE 47. BANDWIDTH LIMITED SQUAREWAVE AND ITS SPECTRUM.

DOCID: 3726612

wave and using a complex wave is that in the first instance the modulating wave is a single frequency and in the second instance the modulating wave is a band of frequencies. When a band of frequencies constitutes a modulating wave, each frequency in the spectrum of the modulating wave produces its own pair of side frequencies which straddle the carrier frequency. If, for example, a periodic wave train made up of bandwidth-limited square waves such as that shown in Figure 47A were to be the modulating wave, each spectral component shown in Figure 47B would generate a pair of side frequencies whose amplitude would be proportional to the component causing their generation. The comparative spectra of the modulating wave and the modulated wave are shown in Figure 48. One hundred per cent square-wave modulation is assumed (which of course constitutes on-off keying). Several observations of the spectra in Figure 48 are in order.

1. The Upper Sideband

The upper sideband is an exact replica of the spectrum of the modulating wave. For comparison, the spacing between spectral components is the same in the upper sideband as in the modulating wave. The order in which the components appear is also the same, that is, the side frequency representing the fundamental of the modulating wave appears first above the carrier, the side frequency representing the third harmonic in the modulating wave appears next, and finally the side frequency representing the fifth harmonic. The relative amplitude of the components of the modulating wave and those components of the upper sideband are also identical.

Again, the spacing of the side frequency representing the fundamental component of the modulating wave appears as far above the carrier frequency as the fundamental frequency appears above zero-frequency (dc). Since the spacing, order and relative amplitude of all components of the upper sideband are identical to those same relationships among the harmonics in the modulating wave, it can be readily concluded that both the modulating wave and the upper sideband contain exactly the same bandwidth. Considering these similarities it is perfectly appropriate and proper to say that the information contained in the modulating wave has been translated to a higher part of the frequency spectrum.

Apart from the fact that the upper sideband occurs in a different part of the spectrum than the modulating wave, the only other significant difference is that the components of the modulating wave have given only half of their amplitude to the frequency components of the upper sideband.

DOCID: 3726612

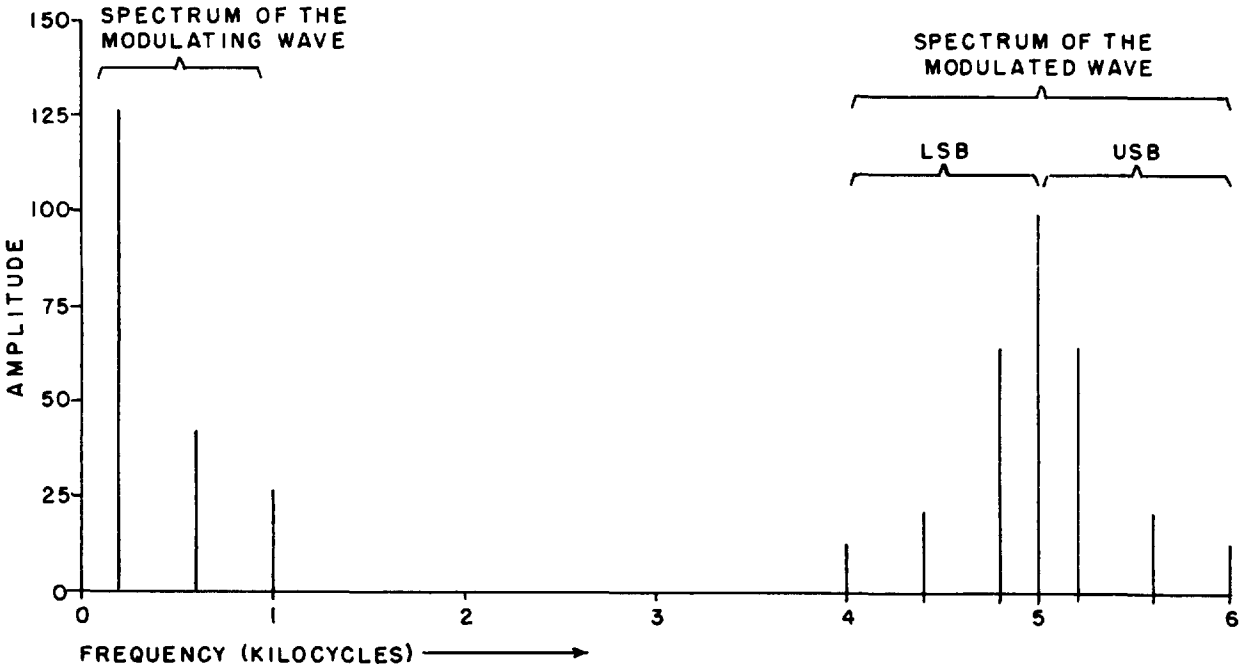


FIGURE 48. COMPARATIVE SPECTRA OF MODULATING AND MODULATED WAVES.

2. The Lower Sideband

The lower sideband is an inverted replica of the spectrum of the modulating wave. Or as an alternative description, one can say that the lower sideband is a mirror image of the upper sideband. This fact makes everything that was said about the upper sideband true of the lower sideband also, except that the order of components is reversed. The reversal simply means that the side frequency representing the fundamental component of the modulating wave appears last (i.e., next below the carrier), the side frequency representing the third harmonic in the modulating wave appears next to last (i.e., second below the carrier) and finally the side frequency representing the fifth harmonic appears lowest of all.

Despite the inversion, the lower sideband is also a case of information contained in the modulating wave being translated to a higher part of the spectrum. Like the upper sideband it contains all of the same information that the modulating wave contains and requires the same bandwidth. As to amplitude, the components of the modulating wave have given the other half of their amplitude to the frequency components of the lower sideband.

DOCID: 3726612

V. BANDWIDTH CALCULATIONS

Circuit bandwidth and signal bandwidth have been mentioned and discussed a number of times in this and in the preceding lessons. Signal bandwidth is that bandwidth which is bounded by the upper and lower spectrum limits of the signal in question. Circuit bandwidth, on the other hand, is the equivalent equipment characteristic and is generally considered to be a frequency band where upper and lower limits occur where the attenuation is 3 db below the average level of the band. Circuit bandwidth determines the fidelity with which a signal will be reproduced.

A. Intelligence Bandwidth and Transmission Bandwidth

Prior to modulation, signal bandwidth consists solely of the spectrum bandwidth of the modulating wave and may be referred to either as intelligence bandwidth or recording bandwidth depending upon the application.

After modulation, signal bandwidth comprises the spectrum bandwidth of the modulated wave. This spectrum includes the carrier and both sidebands. This may be referred to as the sideband bandwidth or the transmission bandwidth. (Of course if the former term is used, one must specify whether the bandwidth of one sideband or total sideband bandwidth is intended.) For amplitude modulation the transmission bandwidth is twice as wide as the intelligence or recording bandwidth. (This assumes that the intelligence bandwidth goes all the way down to dc.)

B. Bandwidth Considerations for Voice

A typical voice wave and its spectrum were shown in Figure 49. The frequencies involved in normal speech vary from one individual to the next but the average speaking voice of the adult male has a fundamental frequency of about 128 cycles per second which corresponds roughly to C below middle C on the piano. A deep-voiced male may have a low of about 90 cycles per second. The average speaking voice of the adult female has a fundamental tone of about 256 cycles per second--a frequency twice that of the average male voice and corresponds roughly to middle C on the piano. A shrill-voiced soprano may have a fundamental pitch of about 300 cycles per second. Other factors extend the range of the speaking voice spectrum. The first of these factors is the blending of overtones or harmonics up to as far as 6 kilocycles per second and the second factor is the sound produced by certain consonants such as the hiss of the letter "S" which may go as high as 10 kilocycles per second. The singing voice has an even greater range going as low as 60 cycles per second for a deep bass and as high as 1300 cycles for a high soprano with harmonics going up to 10 kilocycles per second.

DOCID: 3726612

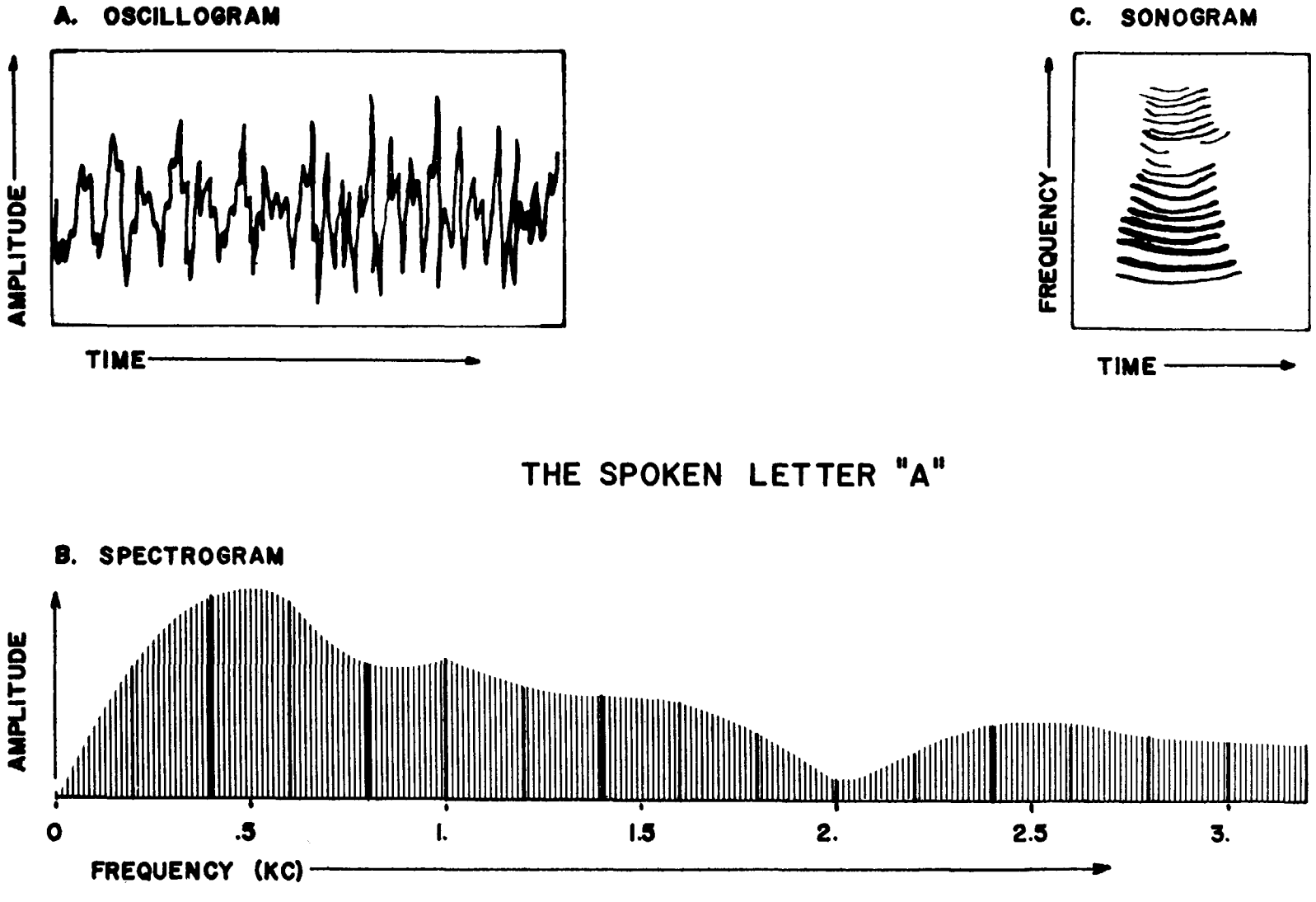


FIGURE 49. THREE VIEWS OF A VOICE WAVE

DOCID: 3726612

1. Recording Bandwidth

The information bandwidth required for voice depends upon the desired quality or fidelity. Entertainment functions require the greatest bandwidth because they must not only accommodate the spectrum of speech and vocal music but instrumental accompaniment, as well. For example, standard A-M broadcast permits an information bandwidth of 100 to 5,000 cycles per second and A-M stations licensed to broadcast high quality music programs (high fidelity is not possible on A-M with present legally established bandwidth) permit an information bandwidth of 35 to 8,000 cycles per second. In either of these instances the allowed bandwidth is considerably less than the audible range of frequencies which is considered to extend from 20 to 20,000 cycles per second. The effect of such bandwidth limiting is shown in Figure 50 where the information bandwidth of a speech wave is reduced by circuit bandwidth from 4 kilocycles to 1 kilocycle.

Pure communications functions use the least bandwidth since they are concerned with speech only. The most important frequencies of speech fall almost entirely below three kilocycles. Most of the energy within this three kilocycle band is concentrated in the frequencies below one kilocycle (see Figure 49) but it happens that the frequencies having the most energy contribute the least to the intelligibility of words. The tones above one kilocycle possess very little of the total energy of a 3 kc speech spectrum (about 17%) but they are the most important to the intelligibility providing about 86% of perfect articulation.

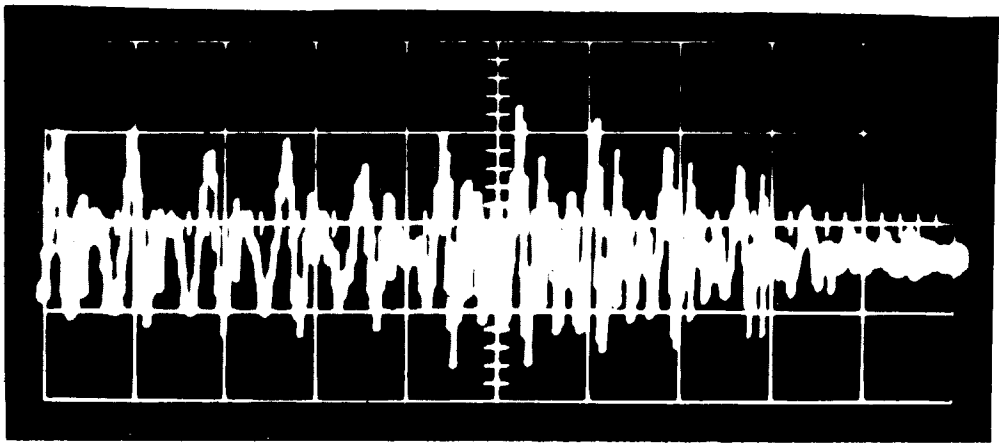
It is found that the minimum band for transmission of intelligible and identifiable speech is 500 to 2,000 cycles per second. The general <u>average</u> band in use is 250 to 2,750 cycles per second but the <u>best</u> current practice is to consider 100 to 3,500 cycles per second as the voice range for pure communications. The table below lists some typical values.

Standard Voice	300 to 3,000 cycles.
High Quality Voice	- to 4,000 cycles.
Standard Broadcast	100 to 5,000 cycles.

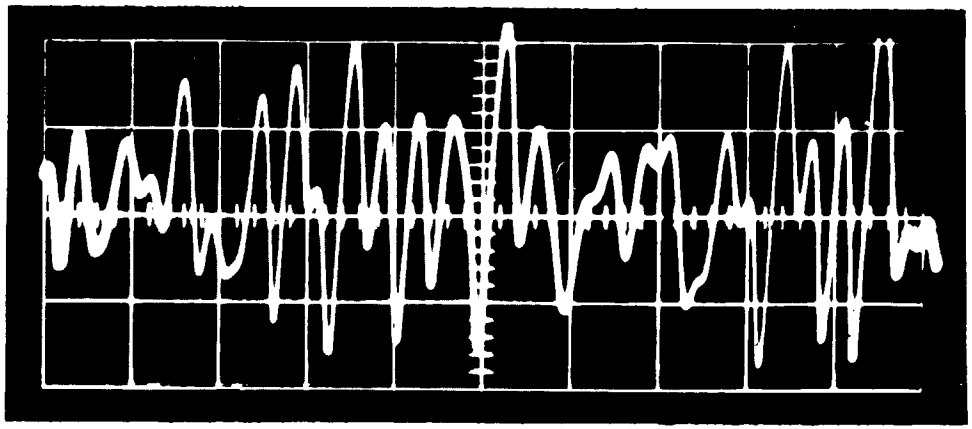
2. Transmission Bandwidth

Since in A-M each frequency in the spectrum of an information

DOCID: 3726612



(A) SPEECH WAVEFORM WITH 4KC BANDWIDTH.



(B) SPEECH WAVEFORM WITH IKC BANDWIDTH.

FIGURE 50.

wave produces one pair of sideband components there will be twice as many frequencies in the sidebands as there are in the modulating wave. The highest side frequency generated will be the sum of the carrier frequency and the highest frequency in the information wave. This is the upper limit of the upper sideband. The lowest side frequency generated will be the difference between the carrier and the highest frequency in the information wave. This is the lower limit of the lower sideband. The band existing between the outer limits of the two sidebands constitutes the transmission bandwidth. Since the band required by the sidebands is the carrier plus and minus the highest modulating frequency, the transmission bandwidth is obviously double the recording bandwidth. If the voice frequency is considered to extend to 2.75 kc the transmission bandwidth required by the sidebands will be double this value or 5.5 kc. If the upper limit is 3.5 kc then the transmission bandwidth is 7 kc. A formula can be written that derives the transmission bandwidth from the recording bandwidth and, of course, gives an indication of the minimum circuit bandwidth required for transmission or reception. This formula is,

$$BW = 2 F_m$$

where F_m is the highest modulating frequency. In consideration of this, Figure 51 gives typical voice A-M transmission spectra based on the table previously given.

C. Bandwidth Considerations for Rectangular Pulses

Pulses, particularly rectangular pulses were discussed at length in a previous section--Pulse Analysis. Rectangular pulses forming square waves were found to contain an infinite number of odd harmonic multiples of the reciprocal of the wave period. Other periodic rectangular waves were found to contain both even and odd harmonics, the ratio of pulse duration to pulse period always determining which harmonics will drop out. It was also found that decreasing the pulse period spaces the harmonics farther apart and decreasing the pulse duration increases the bandwidth of the significant pulse spectrum. All of these characteristics were shown to be very closely related to a curve called a (sineX)/X curve which defines the envelope of the spectrum of any given rectangular pulse. This curve was then found to be applicable to the bandwidth requirements of all rectangular pulses even when they are aperiodic or transient.

1. Recording Bandwidth

Up to this point all that has been taught concerning the

DOCID: 3726612

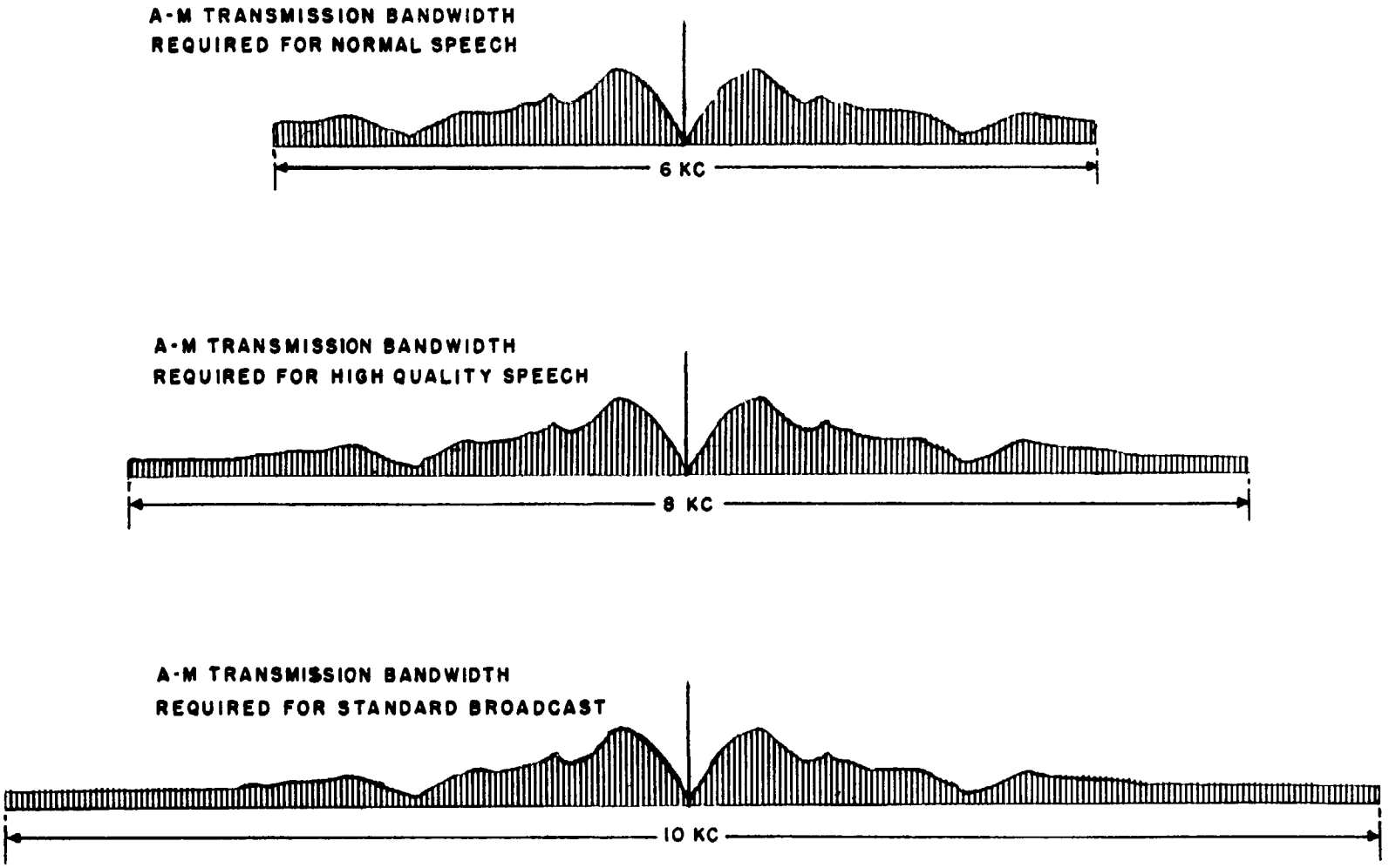


FIGURE 51. A–M TRANSMISSION BANDWIDTHS FOR SPEECH AND MUSIC.

DOCID: 3726612

spectrum of pulses has been in regard to the information bandwidth or recording bandwidth. The bandwidth in all actual situations must always be something considerably less than the theoretical bandwidth of a perfect pulse. The information bandwidth required for a pulse in any particular situation depends upon the desired quality or fidelity.

Generally the inclusion of the first and third harmonics is the base minimum bandwidth for a periodic squarewave. In terms of the (sine X) /X curve this bandwidth extends to midway between the first and second (sine X) /X dropouts. Where the first and third harmonics provide an adequate spectrum for a squarewave, a little extra tolerance in bandwidth is allowed to include the second (sine X)/X dropout. This extends the bandwidth to the fourth harmonic (although the fourth harmonic for a squarewave has zero amplitude). The additional tolerance provides for unattenuated passage of the required spectrum.

The bandwidth that will, without detrimental loss of fidelity, pass a squarewave that has some specified pulse duration can, with equal or greater fidelity, pass any rectangular pulse (whether periodic or aperiodic) that has equal or greater pulse duration (see Figure 52). Since transient pulses do not have discrete spectral components but rather a continuous energy distribution, one cannot think of the bandwidth in terms of a certain number of harmonics in their case, although the bandwidth can be considered in terms of the harmonics of an equivalent squarewave. When describing pulse spectrum bandwidth as extending to a certain position along a (sine X)/X curve, one is actually basing bandwidth on an equivalent squarewave (i.e., a squarewave whose pulse duration is equal to the duration of the transient rectangular pulse). Any system that meets the minimum requirement for a squarewave of specified pulse duration will more than accommodate the minimum requirement for any other rectangular pulse whose duration is not less than that of the squarewave.

2. Transmission Bandwidth

When a pulse is used as a modulating wave, every frequency contained in the pulse will produce a pair of side frequencies in the same manner as when a voice wave or when a single sinewave is used as a modulating wave. In any case, the aggregate of the side frequencies produced are the sidebands. The upper sideband will

DOCID: 3726612

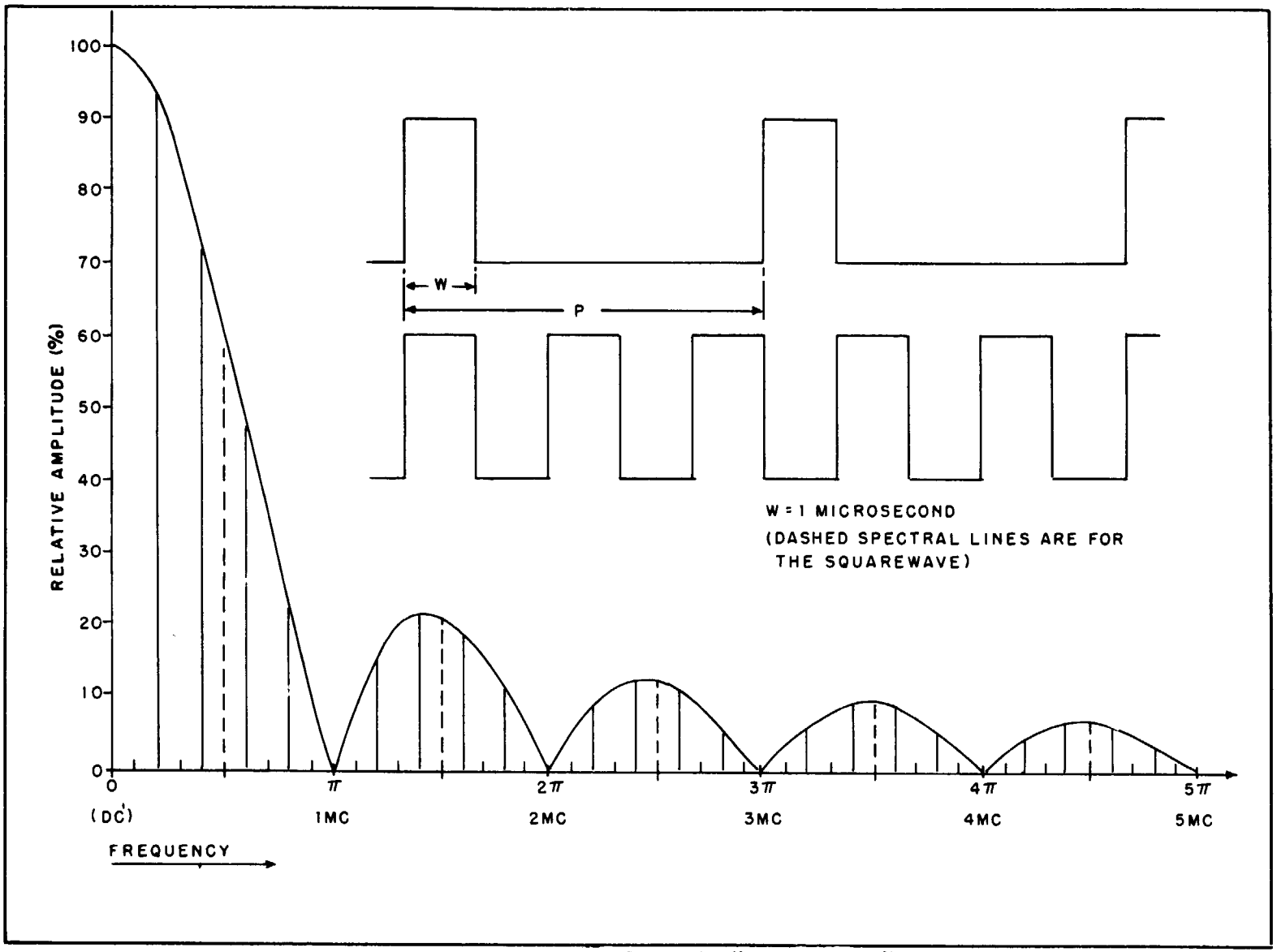


FIGURE 52. SPECTRUM OF A RECTANGULAR WAVE AND AN "EQUIVALENT" SQUAREWAVE COMPARED.

DOCID: 3726612

be an exact image of the spectrum of the modulating wave (but translated above the carrier frequency) and the lower sideband appearing below the carrier frequency will be another image of the spectrum of the modulating wave but with the spectral components in reverse order. Another approach is to say that the upper sideband is an exact image of the modulating spectrum and the lower sideband is a mirror image of the upper sideband. This being true, the spectrum envelope for A-M modulation produced by rectangular waves will be two (sine X)/X curves placed back-to-back with the carrier frequency appearing at the inflection point where the two curves join. Such an envelope appears in Figure 53.

The total bandwidth required by the sidebands (upper and lower together) is equal to the carrier plus and minus the highest modulating frequency. This is double the recording bandwidth just as it was for waves previously described. Thus, if the recording bandwidth for a given rectangular pulse extends to the third harmonic of the equivalent squarewave the transmission bandwidth will be equal to double the frequency of the third harmonic. If extra tolerance to the fourth harmonic is required in the recording bandwidth (as described previously) then the transmission bandwidth must be based upon the fourth harmonic. Whatever frequency is considered as the upper limit for the recording bandwidth, twice this value must be considered as the transmission bandwidth.

Only narrow signal bandwidths have been considered for rectangular pulses. There is good reason not to strive to transmit idealized rectangular waves. When part of a modulating wave, the high order harmonics that are requisite to the very abrupt rise and decay of highly idealized rectangular pulses produce high order side frequencies. These side frequencies may fall outside the assigned channel bandwidth. A typical instance where this problem exists is with ON-OFF keying. These high order side frequencies are usually weak but not so weak as to escape detection or to not cause interference that goes by the name "key clicks."

To prevent interference by key clicks, it is necessary to suppress the higher order harmonics thus preventing the generation of high order sidebands that exceed the channel bandwidth assignment. By this suppression both the transmitting and recording bandwidths are reduced but the wave is distorted and loses much of its rectangularity. To keep the bandwidth as narrow as possible a necessary compromise is made between elimination of

DOCID: 3726612

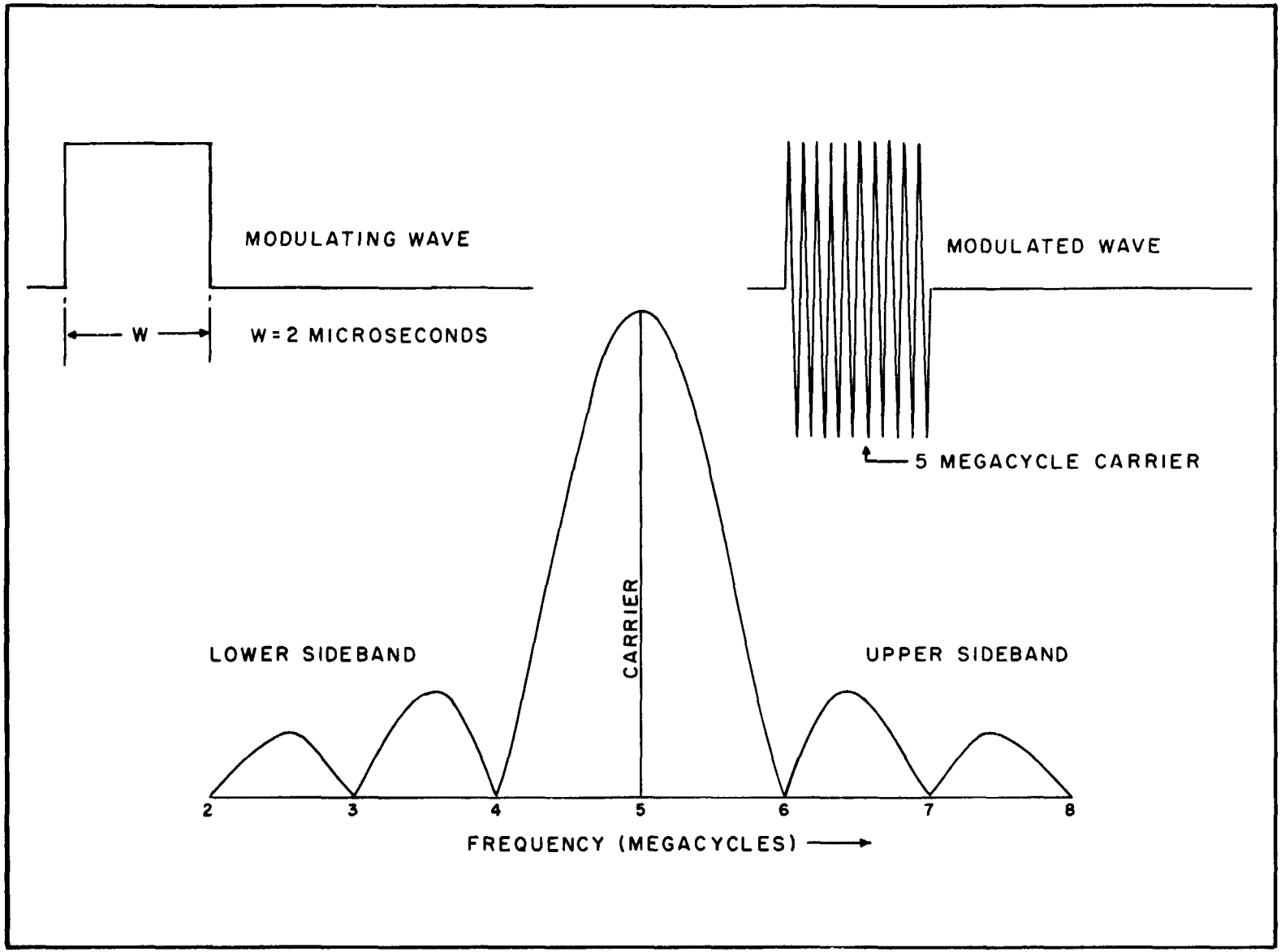


FIGURE 53. A-M TRANSMISSION SPECTRUM FOR A RECTANGULAR PULSE.

DOCID: 3726612

interference (reduction of signal bandwidth) and preservation of fidelity. The result of the compromise is that for either a square-wave, or the squarewave equivalent of some other rectangular pulse, all harmonics above the third are suppressed when the transmission is not subject to fading. When the transmission is subject to fading all harmonics above the fifth are suppressed.

CONCLUSION

Amplitude modulation is one means of encoding an information wave. The process takes each of the component frequencies of the information wave and translates it to new frequencies above and below the carrier frequency. The intelligence deviations of amplitude and period in the modulating wave are thereby transferred to the newly created sidebands. As a result, the sidebands duplicate all of the information in the modulating wave. The carrier itself contains none of the information and goes through the modulation process apparently unaffected. The bandwidth required to pass both sidebands is exactly double that which is needed to pass the spectrum of the original intelligence wave (provided the assumption is made that the original intelligence spectrum contains a dc component).

VI. PRACTICAL EXERCISES

1. The bandwidth for an A-M modulated wave is:

   a. one half the intelligence bandwidth
   b. the carrier plus the third harmonic
   c. double the intelligence bandwidth
   d. the upper sideband plus the third harmonic

2. For 100% modulation, the total power in the A-M sidebands is equal to:

   a. 1/3 carrier power
   b. 1/2 times carrier power
   c. 1 1/2 times carrier power
   d. 2/3 carrier power

3. In A-M modulation:

   a. the peak carrier amplitude is actually varied
   b. the carrier frequency is caused to vary
   c. the peak carrier amplitude only appears to vary
   d. the modulated wave contains less power than the carrier

4. A-M sidebands are:

   a. created in pairs
   b. useless and should be suppressed
   c. always higher in frequency than the carrier
   d. always lower in frequency than the carrier

5. The A-M modulated waveform is:

   a. the sum of the carrier and sidebands
   b. the sum of the sidebands only
   c. a carrier whose amplitude actually changes
   d. comparable to the carrier and modulating wave added together

6. In amplitude modulation, the carrier is:

   a. normally lower in frequency than the modulating frequency
   b. normally higher in frequency than the modulating frequency
   c. normally a frequency equal to the sum of the sideband frequencies
   d. normally a frequency equal to the difference of the sideband frequencies

7. In an A-M modulated wave, all of the intelligence is in:

   a. the carrier
   b. the upper sideband only
   c. the lower sideband only
   d. both sidebands

8. The bandwidth required to receive the carrier is:

   a. zero
   b. double the highest modulating frequency
   c. double the carrier frequency
   d. must be determined by the (sin x)/x curve

9. In A-M there is (more power) (the same amount of power) (less power) in the modulated wave than in the unmodulated wave.

10. In terms of percent, what is the highest degree of modulation that can be used in an A-M system?

DOCID: 3726612

11. For a train of squarewaves having a 30 millisecond pulse duration and used to on-off key a one megacycle carrier find the following:

    a. PRF

    b. third harmonic

    c. fifth harmonic

12. What is the theoretical maximum intelligence bandwidth of the wave in question 11?

13. Allowing no tolerance, what is the minimum practical intelligence bandwidth for the wave in question 11?

14. Allowing some tolerance, what minimum transmission bandwidth would be used for the wave in question 11?

15. What will be the lowest and highest actual frequencies in the spectrum of the modulated wave in question 14?

16. A one megacycle carrier is A-M modulated with two sinewaves of 500 and 1500 cycles per second, respectively. What transmission bandwidth is required?

    What frequencies compose the modulated wave?

17. What is the practical information bandwidth of a one microsecond pulse? (Considering the fourth harmonic of an equivalent squarewave)

DOCID: 3726612

18. What is the transmission bandwidth for the pulse in problem 17?

19. The pulse duration is one microsecond. What is the fundamental frequency of an equivalent squarewave?

What is the minimum practical intelligence bandwidth?

20. Assume a frequency of 2.5 kilocycles. 360 microseconds after the beginning of a cycle, what will be the value of θ in radians? In degrees?

DOCID: 3726612

APPENDIX

The equation of a sinewave is given as,

$$a = A \sin \theta \tag{1}$$

where θ is an angle that varies as a function of time and represents any instantaneous angle of a rotating vector. Upon further analysis, however, θ may be shown to include an angle increasing uniformly with time (2π ft, when expressed in radians, and may then be abbreviated, ωt), and another angle, ϕ (the Greek letter phi) which is superimposed on the uniformly increasing angle so as to add to, or take from the total angle. With this interpretation of θ, equation 1 becomes,

$$a = A \sin (\omega t + \phi) \tag{1a}$$

or

$$a = A \sin (2\pi ft + \phi) \tag{1b}$$

In a given situation the phase angle, ϕ, may be equal to zero, in which instance equations 1a and 1b respectively become,

$$a = A \sin (\omega t) \tag{1c}$$

and

$$a = A \sin (2\pi ft) \tag{1d}$$

If two sinewaves are to be considered, one to represent a carrier and one to represent a modulating wave, the two can be distinguished by the use of subscripts after the symbols for amplitude and angle. A subscript (c) can be used for the carrier, and a subscript (m) can be used for the modulating wave, respectively. Thus, if equation 1c were a sinewave carrier, it would be written,

$$a_c = A_c \sin (\omega_c t) \tag{2}$$

and if it were a sinewave modulating wave it would be written,

$$a_m = A_m \sin (\omega_m t) \tag{3}$$

A non-mathematical definition describes modulation as ". . . a process whereby some characteristic of one wave is varied in accordance with another" (53 IRE 11.S1). Following this definition, if the modulating wave of equation 3 were to modulate

DOCID: 3726612

another sinewave represented by equation 1 (repeated here as equation 4 and supplied with subscripts to designate it as a carrier)

$$a_c = A_c \sin \theta_c \tag{4}$$

it is evident that there are only two "characteristics" that could be varied, namely, either the peak amplitude or the angle. Modulating the amplitude results in amplitude modulation and modulating the angle produces angle modulation. Thus, it is shown mathematically that there are only two basic methods of modulating a sinewave.

For amplitude modulation there must be no phase difference between the carrier wave and the complex modulated wave but the instantaneous amplitude of the complex modulated wave must vary about the carrier peak amplitude (A_c) in proportion to the instantaneous amplitude of the modulating wave. The carrier peak amplitude may be looked upon as an average about which the instantaneous peak amplitude of the modulated wave varies.

A function of time is a quantity whose value and variations in value are in some way dependent upon time and variations in time. The instantaneous amplitude of the modulating wave and the carrier wave are both sinusoidal functions of time, that is, they vary sinusoidally as time varies. The amplitude of the modulated wave is a complex function of time derived from the modulating wave and the carrier wave which are functions of time. The amplitude of the modulated wave as a function of time may be represented symbolically--a(t)--which will be read, "amplitude as a function of time."

The fractional change in amplitude due to modulation is expressed by a modulation factor, M_a, which, expressed decimally, may have values anywhere from 0 to 1. The modulation factor is simply a ratio of the peak amplitude variation used, (A_m), to the maximum design variation (A_c). Expressed,

$$M_a = \frac{A_m}{A_c} \tag{5}$$

By transposition of equation 5, A_m can be expressed as the product of A_c and M_a

$$A_m = A_c M_a \tag{6}$$

The maximum instantaneous peak amplitude that the modulated wave can have (i.e., the maximum value of a(t)) is the sum of the peak amplitude of the modulating wave and of the carrier.

$$a(t)max = A_c + A_m \tag{7a}$$

DOCID: 3726612

When A_m is expressed in terms of the modulation factor according to equation 6 this becomes,

$$a(t)max = A_c + A_cM_a \tag{7b}$$

Factoring out A_c, there obtains,

$$a(t)max = A_c(1 + M_a) \tag{7c}$$

The minimum instantaneous peak amplitude that the modulated wave can have occurs when A_m is at its peak negative value. This, of course reverses the sign in equations 7a, b, and c indicating the difference (instead of the sum) of the peak amplitudes of the modulating wave and the carrier.

Equations 7a, b, and c consider only the peak values of the carrier wave and the modulating wave but it must be remembered that the modulating wave is a function of time that varies sinusoidally between the positive and negative extremes of $\pm A_m$ according to equation 3. Thus, there are instantaneous amplitudes for the modulating wave other than just the positive and negative peak values. Each of these instantaneous values of the modulating wave will produce instantaneous peak amplitudes for the modulated wave. These may be expressed by substituting equation 3 for A_m in equations 7a, b, or c. (Parentheses used in equation 3 have been dropped.)

$$a(t) = A_c + A_m \text{ Sin } \omega_m t \tag{8a}$$

$$a(t) = A_c + A_cM_a \sin \omega_m t \tag{8b}$$

$$a(t) = A_c(1 + M_a) \sin \omega_m t \tag{8c}$$

The remaining fact to be realized is that since the carrier is not always at its peak value either since it, like the modulating wave, is also a sinusoidal function of time. When this fact is included in equation 8 all instantaneous amplitudes of the modulated wave are accounted for, not just the peaks. The expression of all variations that contribute to a(t) are included when equation 2 is substituted for A_c in equation 8b.

$$a(t) = A_c \sin \omega_c t + A_cM_a \sin \omega_m t \sin \omega_c t \tag{9}$$

There is a trigonometric identity which states:

$$\sin x \sin y = \frac{\cos (x-y) - \cos (x + y)}{2}$$

DOCID: 3726612

By means of this identity equation 9 can be expanded to point out the existence of side frequencies. If x in the identity is made to represent $\omega_c t$ and y is made to represent $\omega_m t$, then by substitution equation 9 yields,

$$a(t) = A_c \sin \omega_c t + \frac{A_c M_a}{2} \left[\cos (\omega_c t - \omega_m t) - \cos (\omega_c t + \omega_m t) \right] \tag{10a}$$

or

$$a(t) = \underbrace{A_c \sin \omega_c t}_{\text{carrier}} + \underbrace{\frac{A_c M_a}{2} \cos (\omega_c - \omega_m)t}_{\text{lower side frequency}} - \underbrace{\frac{A_c M_a}{2} \cos (\omega_c + \omega_m)t}_{\text{upper side frequency}} \tag{10b}$$

Equation 10b completely defines the components of an amplitude modulated wave when the modulating wave consists of a single frequency or tone. The first part of equation 10 is seen to be the original carrier wave of equation 2 totally unchanged by the modulating process. The middle term is the lower side frequency which is equal to the difference of the carrier and modulating frequencies $(\omega_c - \omega_m)$ and the last term of equation 10 is the upper side frequency which is equal to the sum of the carrier and modulating frequencies $(\omega_c + \omega_m)$. It is also observed from equation 10 that the sideband amplitude, $A_c M_a$, is proportional to the modulation factor, M_a, and that this amplitude is equally distributed to the two sidebands, one-half $(\frac{A_c M_a}{2})$ going to each. (The plus-sign preceding the lower side frequency indicates that the cosine wave is initially positive and the minus-sign preceding upper side frequency indicates that the cosine wave is initially negative (see Figure 43 and 45).)

Power in the modulated wave is the sum of the power contributed by the carrier and that contributed by each of the sidebands, thus,

$$P_m = P_c + P_{lsb} + P_{usb} \tag{11}$$

Earlier in the text this was expressed by the equation,

$$P_m = P_c \left(1 + \frac{M_a^2}{2}\right) \tag{12}$$

Since power is proportional to the square of the amplitude of current or voltage, equation 12 can be easily derived from equations 10 and 11. Carrier amplitude is indicated by A_c and the amplitude of either sideband is indicated by $\frac{A_c M_a}{2}$, therefore the proportionate carrier power is A_c^2 and the proportionate power in either

DOCID: 3726612

sideband is $(\frac{A_c M_a}{2})^2$. By substitution into equation 11, then follows,

$$P_m = A_c^2 + (\frac{A_c M_a}{2})^2_{LSB} + (\frac{A_c M_a}{2})^2_{USB} = A_c^2 + 2(\frac{A_c M_a}{2})^2 \quad (13a)$$

Squaring as indicated,

$$P_m = A_c^2 + \frac{A_c^2 M_a^2}{2} \quad (13b)$$

Factoring out A_c^2,

$$P_m = A_c^2 (1 + \frac{M_a^2}{2}) \quad (13c)$$

Since A_c^2 represents the carrier power, P_c, the latter can be substituted into equation 13c

$$P_m = P_c (1 + \frac{M_a^2}{2}) \quad (12)$$

to yield equation 12.

The first term in the right hand member of equation 13b represents the carrier power, and the second term represents the total sideband power. That the sideband energy is due to the modulating wave has previously been assumed and asserted but no mathematical evidence was offered. If M_a in equation 13b is expressed in terms of A_m and A_c (refer to equation 5) this may be shown

$$P_m = A_c^2 + \frac{A_c 2 M_a^2}{2} \quad (13b)$$

Substituting equation 5 for M_a into equation 13b

$$P_m = A_c^2 + \frac{A_c 2 (\frac{A_m}{A_c})^2}{2} \quad (14a)$$

Squaring and simplifying

$$P_m = A_c^2 + \frac{A_m^2}{2} \quad (14b)$$

Hence equation 14b demonstrates that the sideband power is related to the modulating amplitude in a way similar to that by which carrier power is related to the carrier amplitude.

DOCID: 3726612

CHAPTER 4

SINGLE SIDEBAND

In the preceding chapter, amplitude modulation was described in some detail. A form of amplitude modulation that was not discussed previously because it warrants a treatment of its own is single sideband. Single sideband proves to be one of the most useful of modulation techniques finding wide use in the transmission of voice and teletype. This chapter will survey the subject briefly.

I. TERMS AND DEFINITIONS

Here are a few terms and definitions that are peculiar to single sideband and its applications, although single sideband is in reality a form of amplitude modulation. These definitions are grouped together at the beginning of the text so they can be easily found for reference, but some of these terms together with their definitions may be more easily understood after the material of the text has been developed.

A. Single Sideband (SSB)

In the discussion on amplitude modulation some mention was made of frequency translation, where the frequency of some information waves is, by means of modulation, translated up or down in frequency. Translation of frequency has its widest application in taking low frequency waves, usually in the audio range and incapable of being radiated, and translating them up into the radio frequency range where they can be transmitted through the medium of space to some distant receiver. The translated information wave appears in the modulated wave as pairs of sideband frequencies to represent each of the frequencies of the original intelligence. For A-M a single pair of side frequencies is generated for each modulating frequency. In A-M each of the sidebands contains all of the information that exists in the modulating wave; therefore, the loss of either of the A-M sidebands while retaining the other sideband will not result in the loss of any information. The use of just one of the sidebands while discarding the other is called single sideband. Inasmuch as the single sideband contains only a duplication of the modulating wave (but in another frequency range) this technique may be regarded as a case of pure frequency translation. The Institute of Radio Engineers defines single sideband modulation as, "Modulation whereby the spectrum of the modulating wave is translated in frequency by a specified amount either with or without inversion" (53 IRE 11.S1).

B. Single Sideband - Full Carrier

In addition to removing an unnecessary and unwanted sideband, it is

DOCID: 3726612

sometimes desirable to remove the carrier also--a thing easily done since the carrier contains no intelligence and it does consume power from the transmitter. On the other hand, there may be no real need to either suppress or even reduce the carrier, and the full carrier power may be transmitted along with the one sideband. To distinguish this fact, the expression, single sideband-full carrier, is used.

C. Single Sideband Suppressed Carrier (SSSC)

It has already been pointed out that the carrier can be eliminated without loss of information since a continuous periodic wave contains no intelligence in the first place. The main function has been performed in providing a means of modulation or encoding of the information wave, translating it into the desired portion of the frequency spectrum. Its duty done, the carrier may now be completely subtracted out of the modulated wave in which case it is said to be suppressed. The result is single sideband suppressed carrier.

D. Single Sideband Reduced Carrier

Although the carrier wave is not necessary to the transmission of the single sideband wave after modulation, and neither adds to nor takes from the intelligence in the wave, the carrier or at least another wave identical to it (in frequency and phase) is necessary to the demodulation or recovery of the intelligence at the receiver. For this reason the carrier is sometimes transmitted with the single sideband but at greatly reduced power. This is a compromise between full carrier and suppressed carrier and is called single sideband reduced carrier.

E. Asymmetric (or Vestigial) Sideband

Single sideband transmitters are not always designed or intended to completely suppress the unwanted sideband. This is especially true when the undesired sideband is removed by a filtering process. The result is that some vestige (a small remains) of the unwanted sideband appears in the vicinity of the carrier. Because the full, desired sideband and the greatly attenuated, undesired sideband does not present a balanced spectral pattern around the carrier, this form of single sideband is also called asymmetric sideband. It should be noted that television makes use of this type of transmission. The standard definition of the IRE says of vestigial sideband, "The transmitted portion of the sideband which has been largely suppressed by a transducer having a gradual cut-off in the neighborhood of the carrier frequency, the other sideband being transmitted without much suppression," (56 IRE 9.S1).

F. Double-Single Sideband(DSSB)

Since one sideband is all that is necessary to represent one channel of information giving rise to a single sideband system, then why not use the position that was occupied by the unwanted sideband for another information channel sent

DOCID: 3726612

single sideband? This can be done, utilizing a single carrier, where the upper sideband is produced by one information channel and the lower sideband is produced by a second information channel. To summarize, there are two single sideband channels sharing the same carrier, hence the seemingly self-contradictory term, double-single sideband. This term is abbreviated, DSSB.

II. MATHEMATICAL POSSIBILITY OF A-M SINGLE SIDEBAND

The mathematical possibility of A-M single sideband is not substantially different from that of regular A-M. In fact, the A-M processes are used in producing single sideband. The unwanted sideband (and possibly an unwanted carrier) is simply removed either by cancellation or by filtering.

III. SINGLE SIDEBAND WAVEFORMS

The single sideband waveform at first glance does not look substantially different from any other A-M waveform. Some of these are shown in the oscillograms in Figure 45. The same sinusoidal modulating wave is used for all the waveforms shown. For comparison a standard A-M wave is separated into its component carrier and sidebands and then the resultant SSB wave with full carrier and lower sideband removed is shown.

A. Generation of the A-M SSB Wave

Using sinusoidal modulation so it is possible to speak in terms of single frequencies, an observation of the two waves and their components in Fig. 54 shows that the only difference between the standard A-M wave and the single sideband full carrier wave is the removal of the one sideband. (In actual practice, it must be realized that not a pure tone but rather a complex wave would be used for a modulating wave, so that the sidebands would consist of more than single side frequencies.) The resulting waveform does not appear particularly different from the wave that has both sidebands. The wave is easily recognizable as a kind of amplitude modulation.

B. Analysis of the Single Sideband Waveform

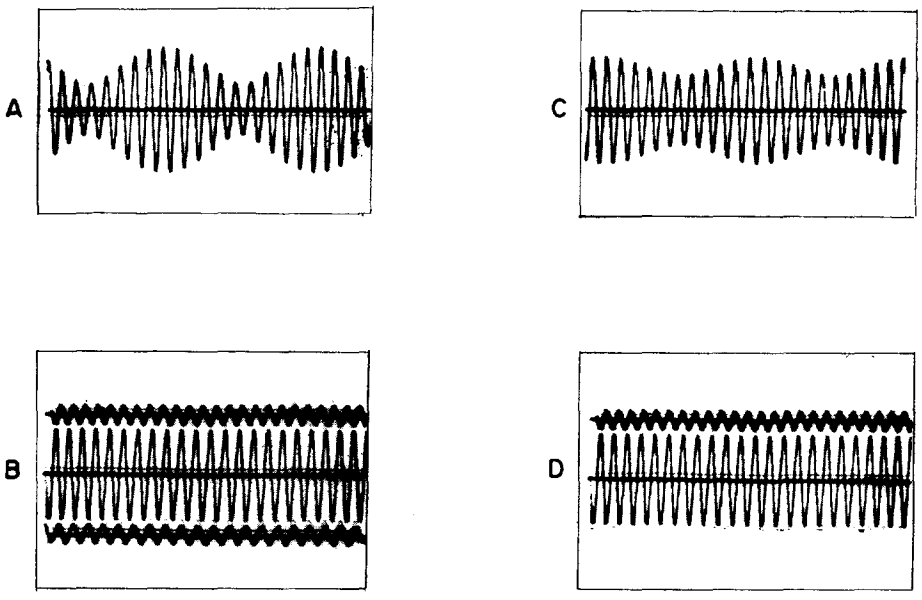
1. Power in the Single Sideband Wave

One of the great advantages of single sideband is the conservation of power. It has been shown that in standard A-M the same information is in both upper and lower sidebands. Even though the information in the lower sideband has an inverted spectrum the information is just as easily recoverable as that in the upper sideband. Since the infor-

DOCID: 3726612

mation from the modulating wave is contained in the sidebands, it is naturally desirable to put as much power into the sidebands as possible. For this reason commerical broadcasting uses the highest depth of modulation possible without running the risk of over modulating. To accomplish this, they use a modulation index of 85% to 95% (100% modulation is not used to allow a margin of safety). Since there is redundant information in the transmitting of both sidebands the question naturally arises, "Why waste power to transmit a sideband that is not needed?" The answer is to use single sideband. The power saved by eliminating one of the sidebands can be added to the power in the other sideband making it approximately double its former value. An examination of the spectrum drawing in Figures 55A and 55B will make it readily apparent what has happened. (One might wonder why single sideband is not used by broadcasting stations if it is so advantageous in terms of efficient use of power. One of the primary reasons that no such service is licensed is that the procedure in proper reception is not as simple as that for standard A-M.)

One further step can be made along the lines of efficient use of



(A) STANDARD A-M WAVE CONSISTING OF THE CARRIER AND TWO SIDE FREQUENCIES SHOWN IN (B) BELOW.

(C) SINGLE SIDE BAND WAVE RESULTING FROM WAVE (A) AFTER LOWER SIDE FREQUENCY IS REMOVED. COMPONENTS OF SSB WAVE ARE SHOWN IN (D) BELOW.

FIGURE 54. STANDARD A–M AND SINGLE SIDEBAND WAVES COMPARED.

DOCID: 3726612

power. The carrier contains none of the intelligence whatever. It is needed, however, to aid in demodulation of the signal at the receiver, and is useful to provide a signal for automatic gain control at the receiver. Where the automatic gain control can be handled by other means, it is possible to replace the carrier in the demodulation function by generating a signal at the carrier frequency by means of a local oscillator in the receiver. If this is done, there is no longer any point in transmitting a carrier that accounts for 2/3 of the power transmitted in a standard A-M signal. The carrier can be reduced or completely suppressed allowing for a tremendous saving in power that can be added to the single sideband. Compare the progressive improvements in power utilization in Figures 55A, 55B and 55C.

2. Frequency Translation

Single sideband with its carrier suppressed is a case of pure frequency translation. If the upper sideband is used each frequency in the modulating wave is increased by an amount equal to the carrier frequency (observe Fig 56). If, on the other hand, the lower sideband is used each frequency in the modulating wave is changed to a frequency equal to the difference between it and the carrier. The upper sideband is said to be normal because it is an exact image of the modulating spectrum but the lower sideband is said to be inverted because it is a mirror image, or inverted image of the modulating spectrum.

IV. THE SINGLE SIDEBAND SPECTRUM

A comparison of Figs. 55C and 56 illustrates the spectrum of single sideband suppressed carrier for a single modulating tone and for a complex wave. In each case it is simply one-half the spectrum that would result from ordinary A-M having the same modulating waves. For further discussion, if needed, it is suggested that the student review the subject of spectrum as covered in the text on A-M.

V. BANDWIDTH CALCULATIONS

In addition to a saving in power and the efficient use of power provided in single sideband, there is also a considerable saving of bandwidth which at times is a very precious item. The bandwidth required is exactly half that needed for similar A-M transmissions. For crowded bands and limited channel space single sideband provides a welcome economy. Since the A-M bandwidth was given as $BW = 2\ F_m$ where F_m was defined as the highest significant frequency in the modulating spectrum, it is only necessary to consider one-half of this value, or $BW = F_m$, for a single sideband transmission.

DOCID: 3726612

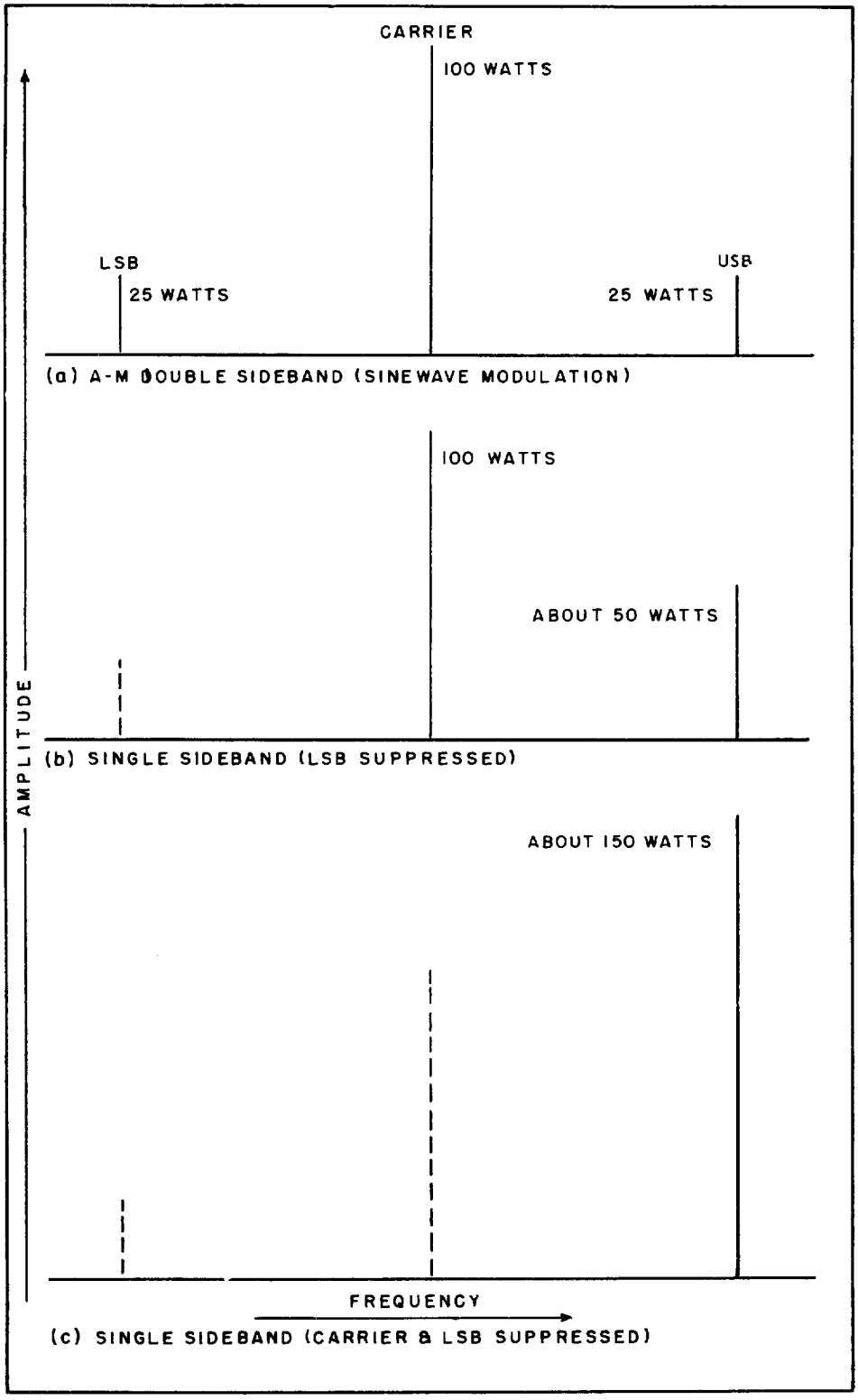


FIGURE 55. A–M DOUBLE-SIDEBAND, SINGLE-SIDEBAND AND SINGLE-SIDEBAND SUPPRESSED CARRIER COMPARED.

DOCID: 3726612

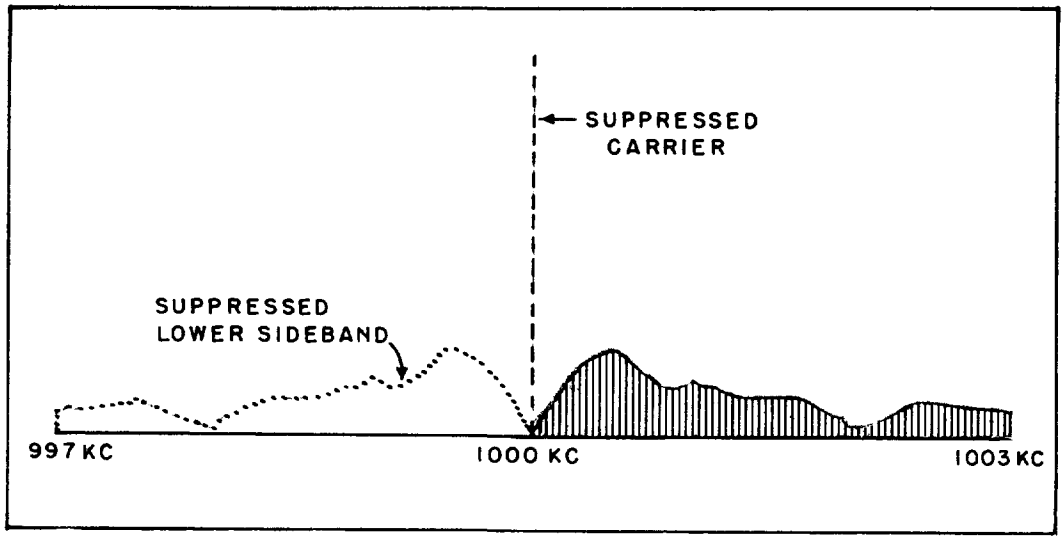


FIGURE 56. TYPICAL SSSC SPECTRUM FOR VOICE.

The equation:

$$BW = F_m$$

includes everything from (and including) the carrier to the highest modulating frequency.

CONCLUSION:

Single sideband, a modified form of A-M, has some decided advantages in its favor. It meets the needs of a limited communication spectrum and doubles the number of available communication channels. Also with the suppression of the unnecessary sideband and reduction of the carrier there is a considerable saving in power that would otherwise be lost. This power can be made available to the remaining sideband for the transmission of useful information. There are disadvantages too, though, the greatest of which is the high degree of frequency stability required at both the sending and receiving ends if the intelligence is to be properly detected. Single sideband and double-single sideband (the latter was not discussed) are not always radiated as such but frequently find place in multiplexing systems where they compose part of a very complex wave used to modulate some other final carrier. This will be the topic of another discussion.

DOCID: 3726612

VI PRACTICAL EXERCISE

1. Single sideband is a form of:

   (a) F-M
   (b) A-M
   (c) P-M
   (d) double sideband suppressed carrier

2. One of the principle advantages of single sideband is that:

   (a) it does not require good frequency stability
   (b) it is free from noise and interference
   (c) it is very conservative of power
   (d) the modulating intelligence exists only in the sidebands

3. Pure frequency translation is most nearly approached by:

   (a) P-M
   (b) A-M (double sideband)
   (c) SSSC
   (d) F-M

4. The bandwidth required for single sideband compares best with:

   (a) double sideband A-M
   (b) the modulating bandwidth
   (c) narrow band F-M
   (d) wideband F-M

5. In a standard A-M signal, how much power is transmitted in this carrier for 100% modulation?

   (a) 1/3 power of the modulated wave
   (b) 1/6 power of the modulated wave
   (c) 2/3 power of the modulated wave
   (d) 1/2 power of the modulated wave

DOCID: 3726612

CHAPTER 5
ANGLE MODULATION

In a previous chapter two basic types of modulation were indicated and amplitude modulation was discussed at length. The other of the two basic types of modulation is angle modulation. This term is unfamiliar to the general public, yet the benefits of one form of angle modulation have been enjoyed by it ever since F-M broadcasting began in 1940. F-M or frequency modulation, however, is not the only form of angle modulation that there is; there is also P-M or phase modulation. Either of these two forms--the first involving an angular rate and the second one involving an angular distance--may be appropriately called angle modulation.

Historically, the theory of angle modulation is much older than the invention. It has been victimized by two controversies. The first of these was whether or not such a method of modulation could have any practical reality. This one has been solved and settled. The second controversy is not a matter of historic record nor even so obvious as the first, but its effects are felt by the uninitiated when they first pursue a study of the subject of angle modulation. This more subtle controversy is embodied in the present failure of engineers and authors to mean the same things by the same terms. The comparison of a random choice of books on the subject will make this latter fact painfully clear.

As the term implies, in angle modulation the phase angle is the wave characteristic that is encoded to carry the frequency and amplitude aspects of an information wave. The two forms of angle modulation, F-M and P-M, will be discussed as separate topics under which the peculiar definitions, possibility of existence, peculiar waveform, spectral structure, and required bandwidth of each will be presented.

I. TERMS AND DEFINITIONS

There are terms and definitions that are peculiar to the forms of angle modulation which will be given here. While it is true that these definitions embody concepts that are not developed until later in the text, they are given at the outset where the student can refer back to them whenever they are needed. The terms, modulating wave, carrier, and modulated wave, however, have a general usage and apply to both angle modulation and amplitude modulation and share similar definitions in both instances. These latter terms, therefore, will not be repeated here but if the student finds it necessary to refresh his memory on these terms and their usage, it is suggested that he refer to the previous lesson where the definitions are given in detail.

A. Frequency Modulation

Amplitude modulation was shown to be a means of encoding an

DOCID: 3726612

information wave by producing a complex wave whose peak amplitude was varied in proportion to the instantaneous amplitude of the information wave. Another means of encoding a wave of information is by producing a complex wave whose frequency is varied in proportion to the instantaneous amplitude of the information wave. Such an encoding is frequency modulation. The result of this encoding or modulation process is a complex modulated wave whose instantaneous frequency is a function of the amplitude of the modulating wave and differs from the frequency of the carrier from instant to instant as the amplitude of the modulating wave varies. The standard definition of frequency modulation describes it as, "Angle modulation of a sinewave carrier in which the instantaneous frequency of the modulated wave differs from the carrier frequency by an amount proportional to the instantaneous value of the modulating wave" (53 IRE 11-S1).

B. Center Frequency

As with amplitude modulation there is an assigned frequency for the carrier wave. Where the modulated wave is to be radiated the frequency and stability of the carrier is set by law. In amplitude modulation this carrier is always present in the complex modulated wave and its frequency is the center of any double sideband A-M spectrum. This, though, is not quite the usage that is given to the term in angle modulation. In frequency modulation, the center frequency is that frequency assigned to the carrier, but during modulation the carrier is not always present in the complex modulated wave, and the instantaneous frequency of the complex modulated wave varies above (i.e., becomes greater than) and below (i.e., becomes less than) the frequency assigned to the carrier. This assigned frequency, then, is the "center" about which the instantaneous frequencies of the modulated wave vary. This conforms with the standard definition which states that the center frequency is ". . . the average frequency of the emitted wave when modulated by a symmetrical signal" (48 IRE 2, 11, 15. S1). The center frequency is a system design constant. Center frequency is abbreviated, f_x.

C. Frequency Deviation

In the foregoing definitions, reference has been made to the variations of the instantaneous frequency of the modulated wave above and below the center frequency. In F-M the shift in frequency is proportional to the amplitude of the modulating wave. A weak or low level modulating wave will, of course, have a small peak amplitude, which, in turn, will produce a small peak frequency variation, and similarly a medium level of peak amplitude will produce an intermediate peak frequency variation. A strong or high level modulating wave whose peak amplitude is the maximum that the modulating system is designed to handle will produce the maximum peak frequency variation. (Note the distinction that exists between the words maximum and peak.) In F-M, the peak frequency shift occurs when the

DOCID: 3726612

modulating wave is going through its peak amplitude. Any of these peak variations from the center frequency is called the frequency deviation (abbreviated f_d); hence, in frequency modulation, frequency deviation is defined as, ". . . the peak difference between the instantaneous frequency of the modulated wave and the carrier frequency" (53 IRE 11. S1). The maximum value of frequency deviation (abbreviated f_D) is a system design constant, and when it is intended that the modulated wave be radiated, its magnitude is established by law. Frequency deviations (f_d) other than the maximum (f_D) are signal characteristics.

D. Frequency Swing

Frequency deviation was defined and shown to be the peak frequency shift to either side of the center frequency. The particular frequency deviation that occurs depends upon the amplitude of the modulating signal. The strongest modulating signal that a given modulating system is designed to accommodate will produce the maximum frequency deviation. The amplitude peaks of the positive alternations of such a modulating signal will produce the maximum positive deviation and the amplitude peaks of the negative alternations will produce the maximum negative deviation. The overall extreme of this excursion of instantaneous frequencies from maximum negative to maximum positive is called the frequency swing. Frequency swing, therefore, is equal to twice the maximum design frequency deviation. The definition for the term frequency swing as applied to frequency modulation is ". . . the difference between the maximum and minimum design values of the instantaneous frequency" (53 IRE 11. S1). In F-M the frequency swing is a system constant and is usually expressed in terms of the maximum frequency deviation as $\pm f_D$.

E. Deviation Ratio

Any modulating system is intended to accommodate some specific band of modulating frequencies. Therefore the lower limit and especially the upper limit of this band are of importance in the design of the equipment. Of the two limits, the upper is most important in any modulating system because it determines the maximum bandwidth requirements. In F-M, the highest modulating frequency, designated f_M, is a system constant. (All other modulating frequencies will be designated f_m.) The ratio of the maximum frequency deviation, f_D (a design constant) to the highest modulating frequency, f_M is another system design constant D, called the deviation ratio. The deviation ratio is strictly an equipment characteristic and is a quantity used to set the circuit bandwidth of a modulating system. The standards of the Institute of Radio Engineers (53 IRE 11. S1) says of deviation ratio, "In a frequency modulation system, the ratio of the maximum design frequency deviation to the maximum design modulating frequency of the system."

DOCID: 3726612

Since the ratio,

$$D = \frac{f_D}{f_M}$$

is an equipment characteristic rather than a signal characteristic not much encounter will be made of it in this text, however, the student should be aware of its existence. In a commercial F-M broadcast system in the United States, f_D is fixed by law at 75 kc and f_M is 15 kc (near the upper limit of human hearing). Therefore the deviation ratio of such an F-M transmitter is, --

$$D = \frac{f_D}{f_M} = \frac{75 \text{ kc}}{15 \text{ kc}} = 5$$

F. Modulation Index

There is a very important signal characteristic that resembles the deviation ratio and under certain circumstances is equal to it. It is the ratio,

$$M_f = \frac{f_d}{f_m}$$

Notice, however, that f_d is any frequency deviation, not necessarily maximum, and f_m is any modulating frequency, not necessarily the highest. In this ratio, signal characteristics are in question, not equipment characteristics. However, there are system restrictions that apply; for example, f_d cannot exceed f_D, and f_m cannot be greater than f_M. Since the modulation index, M_f, is the ratio of any value of f_d to any value of f_m within the limits just set forth, it follows that many different values of M_f are possible: this is quite in contrast to the deviation ratio D which for a given system has but one fixed value.

Before going to the definition of the modulation index, it might be well to consider briefly the factors that make it variable. The first factor that can vary M_f is f_m. Taking commercial F-M broadcast for an example, there are in its modulation band all the frequencies from 50 cycles per second to 15,000 cycles per second. If it is assumed that $f_d = f_D = 75$ kc and f_m is alternately given the values of 15 kc, 1.5 kc and .15 kc, respectively, the modulation indices would be,

DOCID: 3726612

$$M_f = \frac{75 \text{ kc}}{15 \text{ kc}} = 5$$

and

$$M_f = \frac{75 \text{ kc}}{1.5 \text{ kc}} = 50$$

and

$$M_f = \frac{75 \text{ kc}}{.15 \text{ kc}} = 500$$

The second factor that makes M_f variable is f_d. It has been stated that f_d may be equal to, or less than f_D. The value of f_d at any time depends upon the depth of modulation produced by the level (amplitude) of the modulating wave. In an F-M system, since the frequency deviation is proportional to the peak amplitude of the modulating wave yielding large deviations for large peak amplitudes and small deviations for small peak amplitudes, the depth of modulation will be determined by the ratio of a given peak signal amplitude to the maximum peak signal amplitude that the modulating system is designed to handle. The ratio of the deviation produced to the maximum possible deviation (for the system) will be in the same proportion. There will, therefore, be as many values of deviation (f_d) as there are different values of peak amplitude to produce them (but there is only one value of f_D for the system). To illustrate the effect of f_d assume three values of f_d, which, due to different depths of modulation, are equal to f_D, $.8f_D$ and $4f_D$ respectively. If f_D = 75 kc, the values for f_d are 75 kc, 60 kc, and 30 kc. Let it be assumed that $f_m = f_M$ =15 kc. The resulting values of modulation index are,

$$M_f = \frac{75 \text{ kc}}{15 \text{ kc}} = 5$$

$$M_f = \frac{60 \text{ kc}}{15 \text{ kc}} = 4$$

$$M_f = \frac{30 \text{ kc}}{15 \text{ kc}} = 2$$

One observation that should be made now is that the modulation index (M_f, a signal characteristic) and deviation ratio (D, an equipment characteristic) are only equal when $f_d = f_D$ and $f_m = f_M$ whereas the deviation ratio is a design constant used to establish the circuit bandwidth of a modulating system, the modulation index is a signal characteristic used to determine the distribution of sideband energy (spectral bandwidth) around the carrier. (Not all textbooks on the subject make a distinction between M_f and D, depending primarily upon the age of the book.)

DOCID: 3726612

For a definition of modulation index, the standards of the Institute of Radio Engineers are quoted, "In frequency modulation with a sinusoidal modulating wave, the ratio of the frequency deviation to the frequency of the modulating wave" (53 IRE 11. S1). (It should be noted that modulation index is defined for sinusoidal modulating waves only.)

G. Phase Modulation

The other form of angle modulation is phase modulation. This form of angle modulation encodes an information wave by producing a complex wave whose phase is varied in proportion to the instantaneous amplitude of the information wave. This method is called phase modulation. The result of this modulation process is a complex modulated wave whose instantaneous phase (relative to the carrier) is a function of the amplitude of the modulating wave and differs from the phase of the carrier from instant to instant as the amplitude of the modulating wave varies. The definition given by the IRE for phase modulation states that it is, "Angle modulation in which the angle of a sinewave carrier is caused to depart from the carrier angle by an amount proportional to the instantaneous value of the modulating wave" (53 IRE 11. S1).

H. Phase Deviation

In P-M there is a shift in phase relative to the carrier phase that is proportional to the amplitude of the modulating wave. A weak or low level modulating wave will have a small peak amplitude, which in turn, will produce a small peak phase shift, and similarly a medium level of peak amplitude will produce an intermediate peak phase shift. A strong or high level modulating wave whose peak amplitude is the maximum that the modulating system is designed to accommodate will produce the maximum peak phase shift. In P-M, the peak phase shift occurs when the modulating wave is going through its peak amplitude. Any of these peak shifts from the carrier phase angle is called the phase deviation (abbreviated M_p because it is the phase modulation index). Hence, in phase modulation, phase deviation is defined as, "The peak phase difference between the instantaneous angle of the modulated wave and the angle of the sinewave carrier" (53 IRE 11. S1). The maximum value of phase deviation (abbreviated ϕ_D) is a system design constant. Phase deviations M_p other than the maximum (ϕ_D) are signal characteristics.

II. MATHEMATICAL POSSIBILITY OF F-M

The significance of the angle θ in the sinewave equation has been elaborated upon previously in Waveform Analysis and Amplitude Modulation. A discussion of its significance will not be repeated here. In the sinewave equation,

DOCID: 3726612

$$a = A \sin \theta = A \sin (2\pi ft + 0)$$

the possibility of producing a frequency modulation may be realized if a means is contrived whereby the wave is distorted by varying the frequency component of the angle θ in accordance with the amplitude changes of an information wave. As with any other instance where a sinewave is distorted, new frequencies are produced, and wherever there is distortion there must be more than a single pure frequency present.

III. FREQUENCY MODULATION WAVEFORM

The frequency modulated waveform resulting from a sinusoidal modulating wave is shown in Figure 57C. It is the type of waveform that would result if the speed of the rotating rector normally generating a sinewave were increased to greater than its normal angular velocity and then decreased to less than its normal angular velocity.

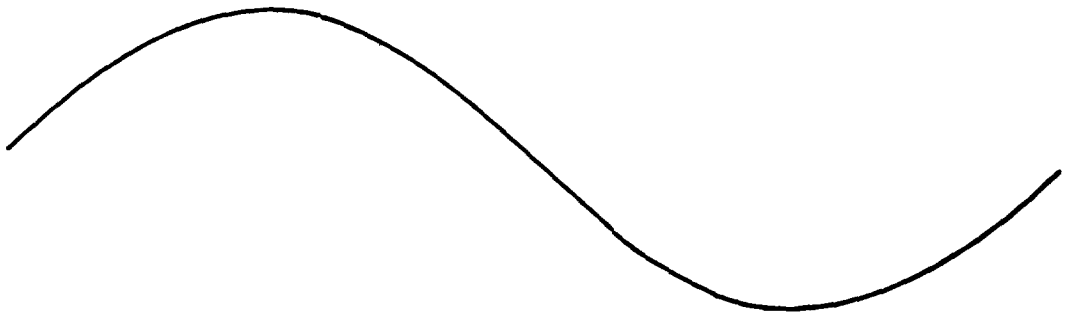
A. Generation of the F-M Wave

If the changes in angular velocity just described were caused by a sinusoidal modulating wave as indicated in Figure 57A, the angular velocity (frequency) would (with respect to its average rate or center frequency) increase during the positive alternations and decrease during the negative alternations as shown by Figure 57C. The effect of the frequency and amplitude of the modulating wave on the frequency deviation is illustrated in Figure 58 where it is seen that the extent of deviation depends strictly on the amplitude of the modulating wave and the rate at which deviation takes place is a function of the frequency of the modulating wave. Thus the two aspects of intelligence in an information wave can be encoded--magnitude of deviation representing the amplitude aspect, and rate of deviation representing the frequency aspect. Figure 58 also illustrates the channel width assigned to a typical F-M station. It will be noted that the channel width consists of double the maximum permissible deviation (i.e., the frequency swing) plus a guard band on each end. The particular values indicated are for commercial F-M broadcast transmitters licensed in the United States.

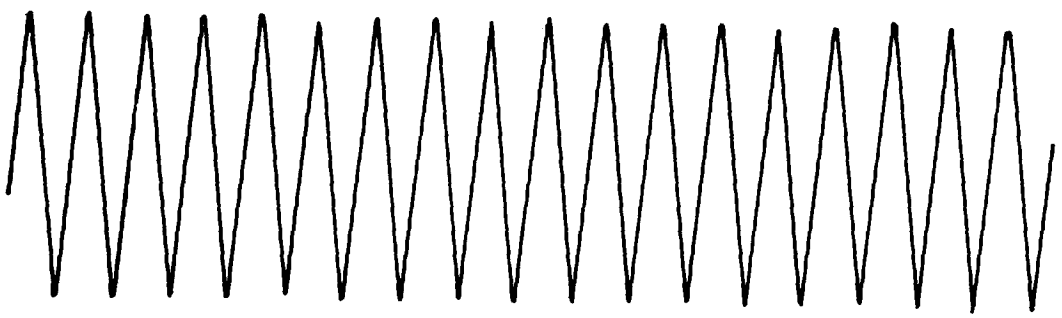
B. Analysis of the F-M Waveform

When a study of amplitude modulation was made, it was found that things do not always happen as they appear to happen. The student should not be greatly surprised if a similar thing is true of F-M. In the A-M situation it appeared that the process of modulation simply caused the amplitude of the carrier to change in accordance with the amplitude of the modulating wave, but analysis showed that the actual carrier remained unchanged while new frequencies, called

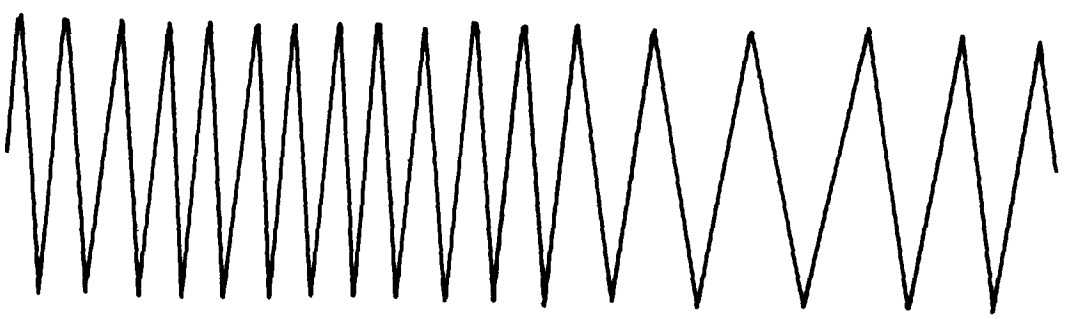
DOCID: 3726612



A. MODULATING WAVE.



B. CARRIER WAVE.



C. FREQUENCY MODULATED WAVE.

FIGURE 57. FREQUENCY MODULATION.

DOCID: 3726612

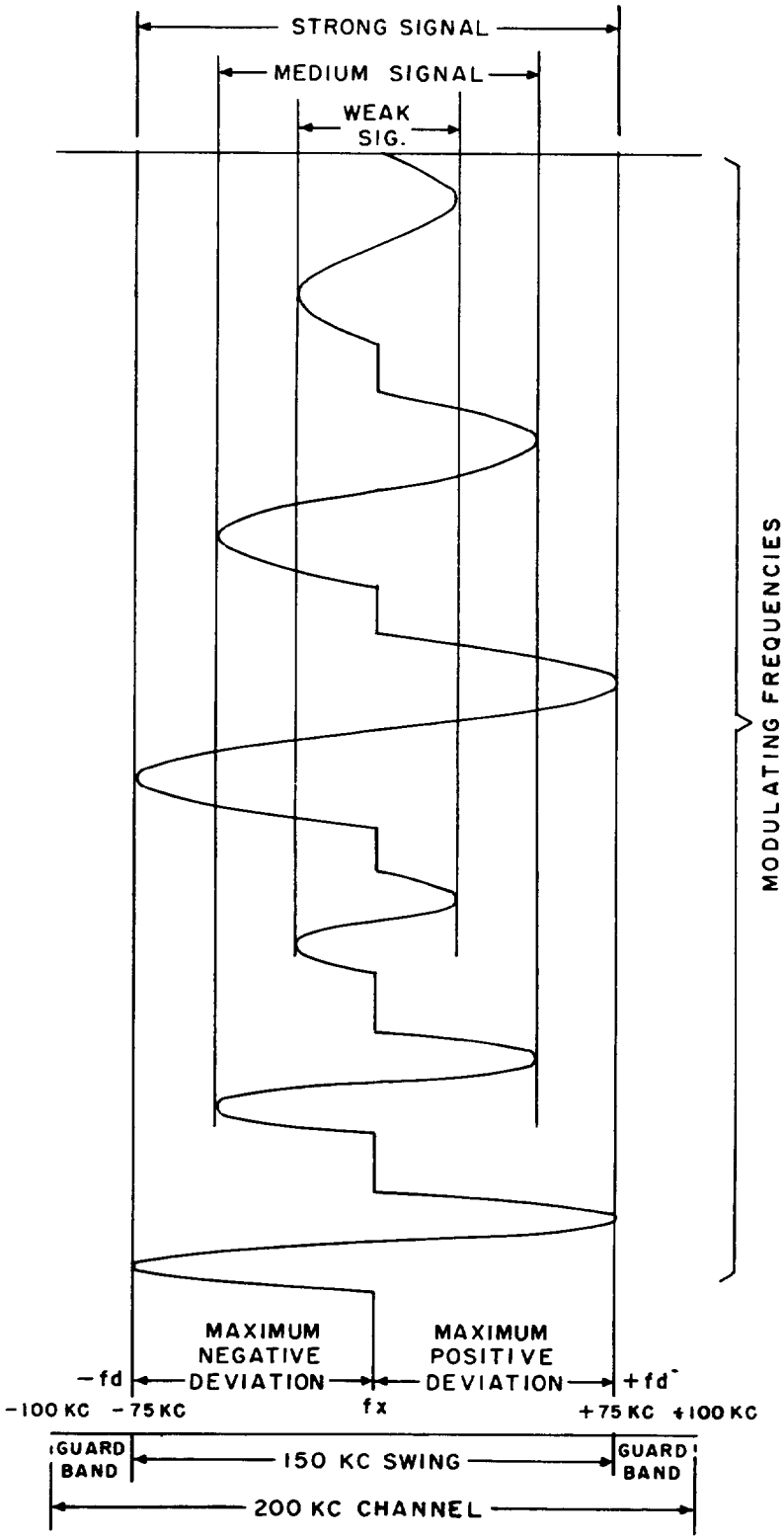


FIGURE 58. THE EFFECT OF MODULATING FREQUENCY & AMPLITUDE ON F-M.

DOCID: 3726612

side frequencies, were produced, which, when added to the carrier, yielded a complex wave having the observed appearance. In F-M, appearance also plays a deceptive role for what appears to be a sinusoidal wave of gradually changing frequency, is actually a complex wave consisting of many frequencies. (Besides, by definition a sinewave cannot change frequency for it is defined as consisting of a single, constant frequency.)

1. Sidebands in F-M

When A-M modulation was produced by a single sinewave modulating wave, there was created a pair of side frequencies, one lower than the carrier frequency by an amount equal to the modulating frequency and another higher than the carrier frequency by an amount equal to the modulating frequency. But this is not so in F-M. When a single sinewave modulating wave produces frequency modulation there is not just one pair of side frequencies created, but theoretically an infinite number of pairs of side frequencies. These straddle the carrier or center frequency in a symmetrical arrangement. The frequencies of these side frequency components are equal to the center frequency plus and minus all harmonics of the modulating frequency. Consequently, these are separated from the carrier and from one another by multiples of the modulating frequency. Figure 59 illustrates a complex F-M modulated wave and the most important sideband pairs that compose it.

2. Power in the F-M Wave

In A-M it was found that there was more power in a modulated wave than in an unmodulated carrier. It was shown that this added power came from the modulating wave. The power in an A-M modulated wave fluctuates as the amplitude of the modulating wave varies. In F-M this is not so. The power in an F-M modulated wave is constant. This is because the modulating wave does not add any power to the wave through the modulation process. Sideband power in F-M does not come from the modulating wave as it does in A-M; instead the sideband power in F-M is extracted from the carrier power and redistributed among the sidebands so that the total power in the resulting complex wave remains the same as that in the unmodulated carrier. Under certain modulating conditions all the power is extracted from the carrier and redistributed among the sidebands in which case the carrier vanishes and the complex wave consists only of sidebands.

DOCID: 3726612

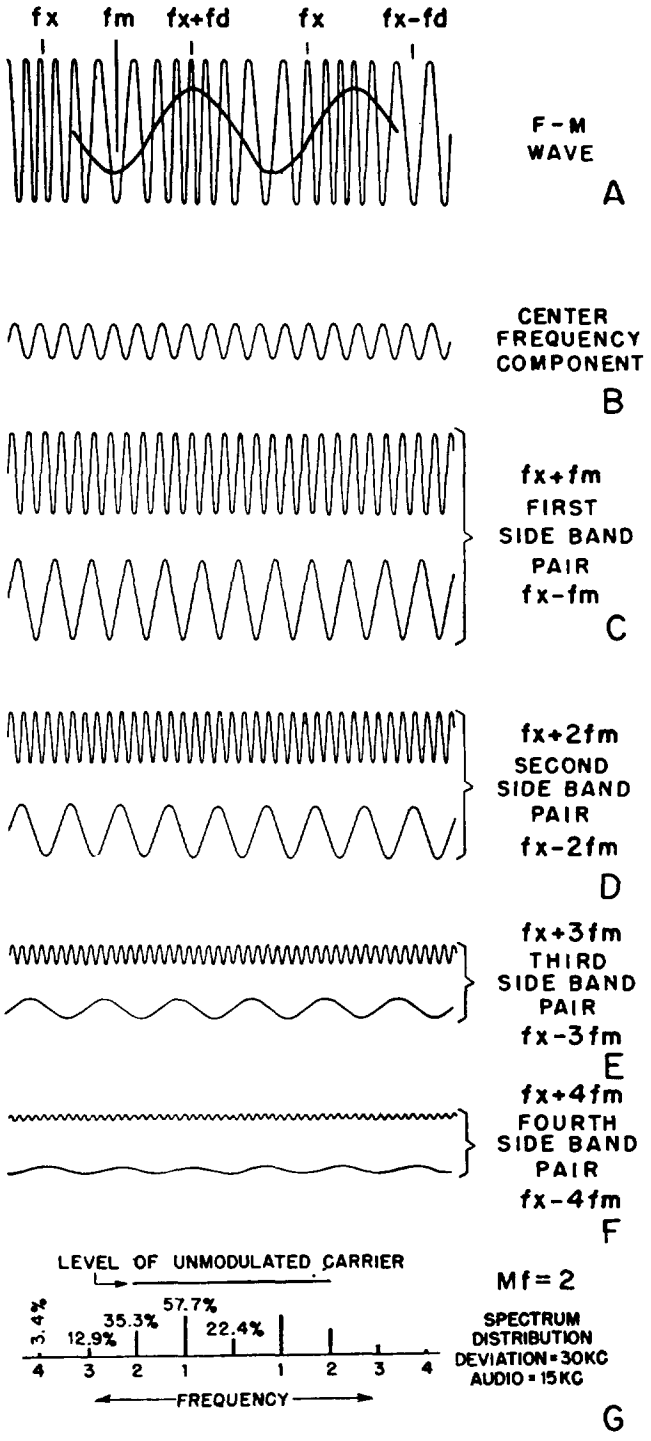


FIGURE 59. F-M CARRIER, SIDEBANDS AND SPECTRUM.

DOCID: 3726612

3. Frequency Translation

F-M accomplishes frequency translation of the modulating information just as A-M does, except that instead of translating each frequency of the modulating wave into two new positions in the frequency spectrum, it translates them into an infinite number of new positions.

IV. THE FREQUENCY MODULATION SPECTRUM

When modulating with a single tone the point has been established that a modulated wave, complex in composition, is generated. Unlike A-M which generates one sideband pair for each modulating frequency, F-M generates an infinite number of sideband pairs for each modulating frequency. The appearance of the F-M spectrum and the significant portion of that spectrum will be discussed in the paragraphs that follow.

A. Spectrum for Single Tone Modulation

Figure 60 illustrates the spectrum generated by a 15 kc sinewave modulating a carrier to produce a frequency deviation of 75 kc. The resultant modulation index is 5.

1. Spacing of Spectral Components.

The separation of the spectral components is seen to be multiples of the modulating frequency. The first sideband order (i.e., pair of side frequencies) straddle the center frequency and are separated from it by an amount equal to the modulating frequency. To elaborate, the upper side frequency of the first order is at $f_x + f_m$ and the lower side frequency of the first order is at $f_x - f_m$, just as in A-M.

The second sideband order is similar to the first sideband order, but this time the frequencies are $f_x + 2 f_m$ and $f_x - 2 f_m$. Similarly the third order is located at $f_x + 3 f_m$ and $f_x - 3 f_m$. The spacing of other sideband orders is similar to these, the multiplier in each case being identical to the number of the order. This system separates all side frequencies by multiples of the modulating frequency.

2. Bessel Function Curves

By now the student has probably noticed an apparent discrepancy between the number of sideband orders displayed in Fig. 60 and the

DOCID: 3726612

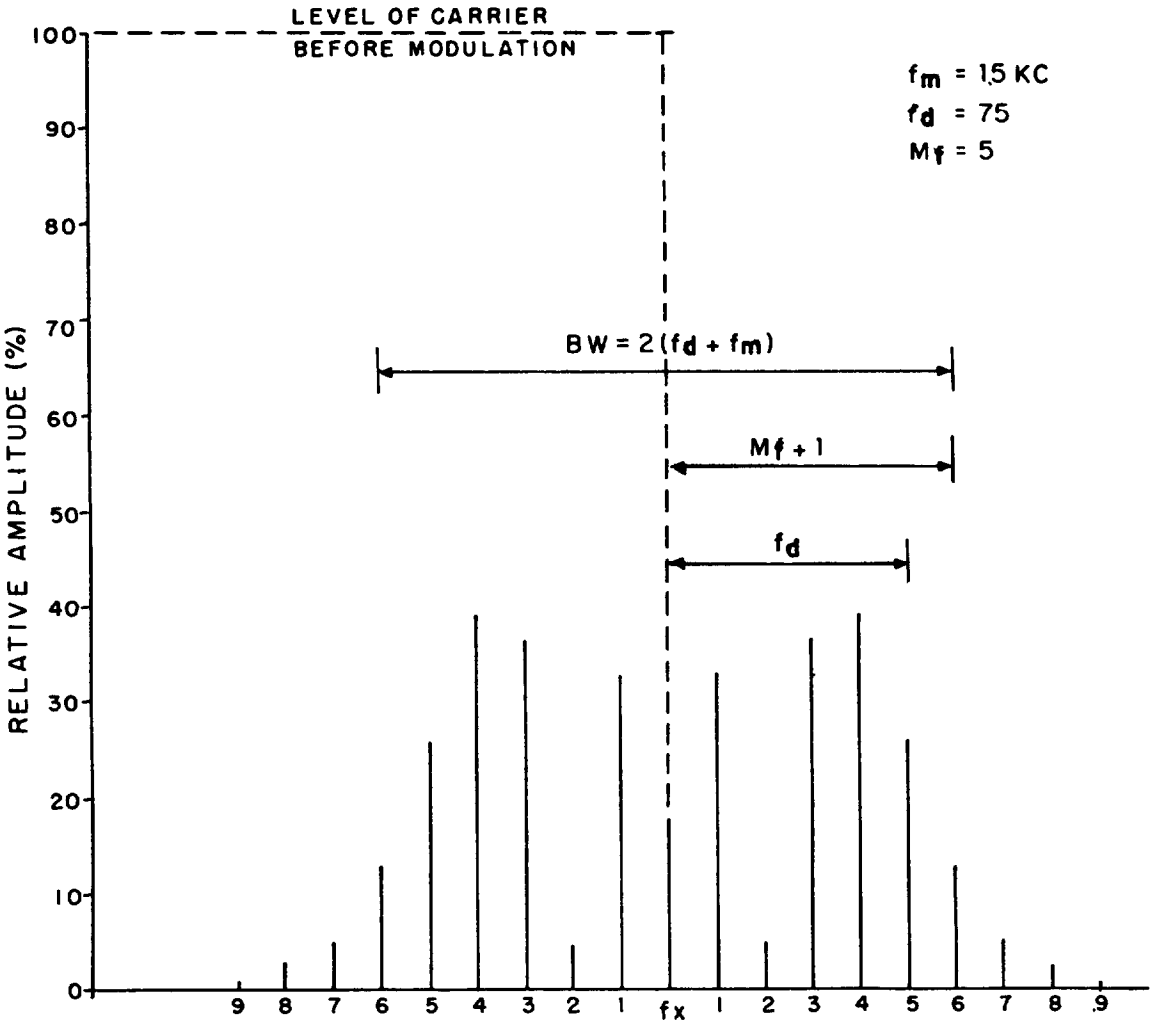


FIGURE 60. A FREQUENCY MODULATION SPECTRUM.

DOCID: 3726612

previous statements that were made about an infinite number of sideband orders being generated. Certainly, nine orders of sidebands are not an infinite (unlimited) number of sidebands! The apparent discrepancy is easily resolved, however. It will be noticed that the carrier (component at the center frequency) is reduced in amplitude to only 17.76% of its unmodulated value, the first order of sidebands is quite large, the second order quite small, then there is a gradual increase to the fourth order of side frequencies, followed by a rapid decay in amplitude for the remaining sideband orders. Beyond the ninth order there is still an unlimited number of side frequencies but none of them have an amplitude great enough to indicate on a drawing of this scale. These all have amplitude of less than 1% of the unmodulated carrier.

The relative amplitude of the side frequencies and of the modulated carrier, all in terms of per cent of the amplitude of the unmodulated carrier, can be found by consulting a chart of Bessel function curves like that shown in Figure 61. Twelve curves are shown in this chart and each is labeled to correspond with the carrier or any of twelve sideband orders. Along the bottom there are listed modulation indices. Where the labeled curves cross any particular modulation index, the amplitude of a curve at that point indicates the relative amplitude of the component designated by its label. The student should find it interesting to check the amplitudes in Figure 60 against those indicated by Figure 61.

The Bessel curve, designating the way that carrier amplitude varies with the changes in modulation index, closely resembles an unmodified (sine X) /X curve but they are <u>not</u> the same. Also, it should be noted that there is no energy at the center frequency at the point where this curve crosses the horizontal axis. These particular points occur at modulation indices of 2.4048, 5.5201, 8.6537, 11.7915 and 14.9309. At these particular modulation indices all of the energy has been extracted from the carrier and redistributed into the sidebands. Such points are often called carrier dropouts.

To summarize, the Bessel function curves and the modulation index of a particular F-M modulated wave are all that is needed to find the relative amplitude of its spectral components, and thus determine which ones are significant enough to include in the required signal bandwidth.

DOCID: 3726612

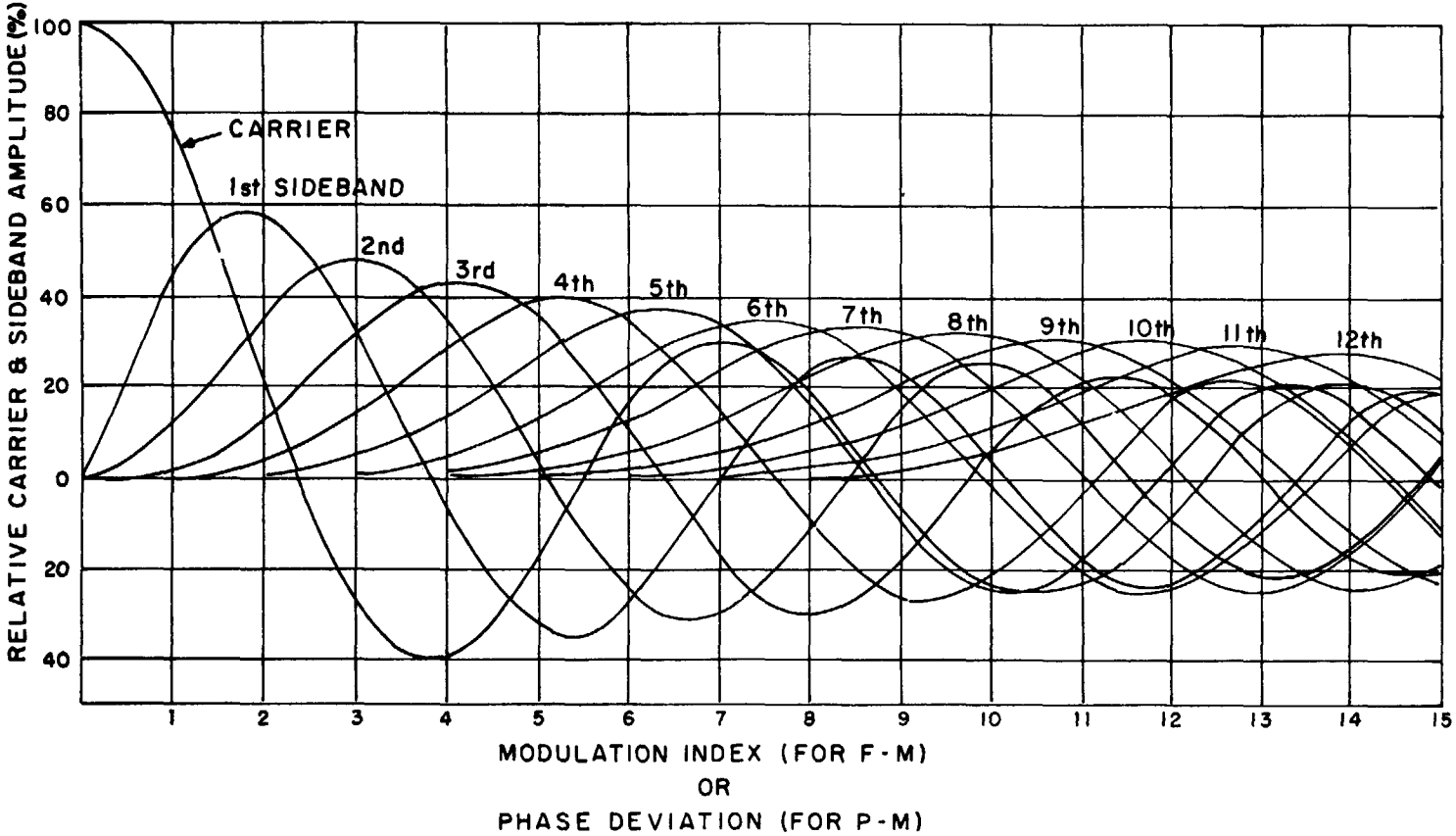


FIGURE 61. BESSEL FUNCTION CURVES.

DOCID: 3726612

B. Spectrum for Multitone Modulation

The spectrum for an F-M modulated wave where the modulating wave is complex results in a rather complex spectrum. First of all, the spectrum will not be symmetrical about the center frequency unless the modulating wave is symmetrical about its horizontal axis. This is because in F-M the upper sideband is associated with the shape of the positive portion of the wave and the lower sideband is associated with the negative portion of the wave. If these two portions do not look alike the resulting spectrum cannot be symmetrical. Figure 62 illustrates a complex wave consisting of two frequencies and the asymmetrical spectrum that results from it. To determine this particular spectrum the modulation index for each of the tones in the modulating wave must be determined separately, the spectrum produced by each tone is computed separately (by means of the Bessel functions), and then the two sets of spectra are added. In some places they will reinforce one another and in other places they will either fully or partially cancel one another. This accounts for the asymmetry of the composite spectrum.

Often, where squarewaves are the modulating wave, as in frequency shift keying (FSK), it is helpful to determine a squarewave modulation index. This is not to be confused with the modulation index defined for sinewaves, although the method of computing it is not dissimilar. The squarewave modulation index is given as follows:

$$\text{Square wave } M_f = \frac{\text{FREQUENCY DEVIATION}}{\text{TRANSMISSION SPEED}}$$

or,

$$= \frac{\text{FREQUENCY DEVIATION}}{\text{PULSE REPETITION FREQUENCY}}$$

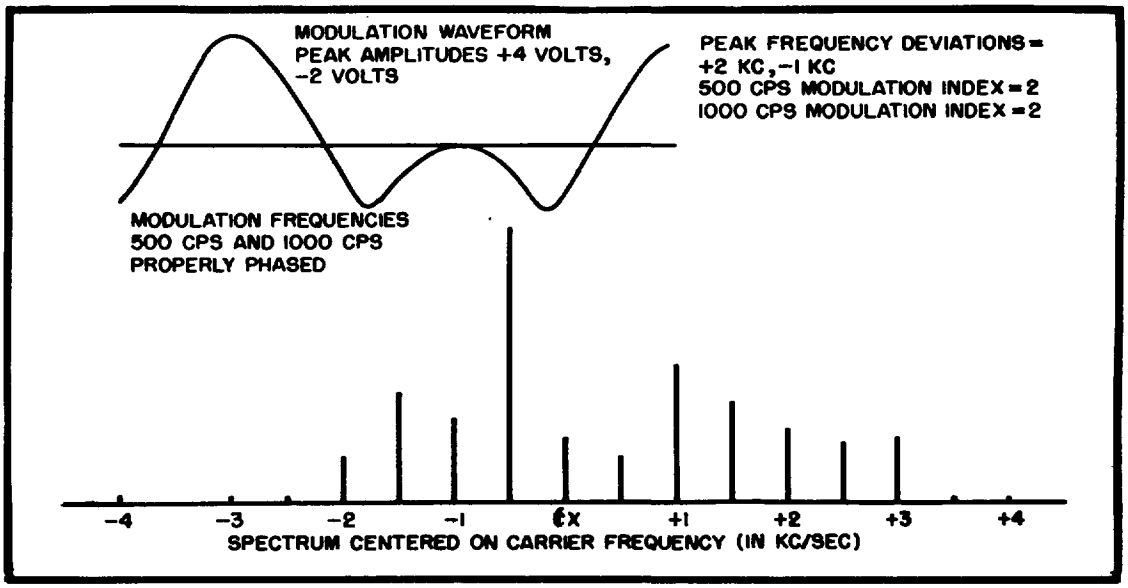


FIGURE 62. FREQUENCY MODULATION BY A COMPLEX WAVE.

DOCID: 3726612

This index will, in most cases, be identical to the system deviation ratio. It may be used in the same manner as the sinewave modulation index to determine the spectrum amplitude relationships. Figure 63 illustrates a squarewave and its F-M spectrum.

V. BANDWIDTH CALCULATIONS

Observation of Figure 60 indicates that the spectral components having the greatest amplitude fall just inside the limits of the frequency deviation, f_d. The number of important sidebands is equal to $M_f + 1$ (i.e., for an M_f of 5, $M_f + 1$ is equal to $5 + 1 = 6$) on each side of the center frequency. This limit of $M_f + 1$ falls just outside the deviation limits but is still within the bandwidth allowed for an F-M channel. It is because of sidebands falling outside the range allowed for frequency deviation that a guard band is provided for every F-M channel. Sidebands falling beyond the $M_f + 1$ limit are always less than 15% of the amplitude of the unmodulated carrier.

The value $M_f + 1$, it must be emphasized, is just the quantity of sideband frequencies on <u>each</u> side of the center frequency that are important. The total number for the signal bandwidth must be <u>twice</u> this value. These facts can be rearranged to give a very useful formula for determining signal transmission bandwidth for F-M. The formula is derived as follows:

Since the sidebands are separated by multiples of the modulating frequency, the frequency of the highest important sideband is given by

$$f_m (M_f + 1) \qquad (1)$$

The over-all bandwidth from the highest to the lowest side frequency whose amplitude is 15% (or greater) of the unmodulated carrier is twice this value or,

$$BW = 2 f_m (M_f + 1) \qquad (2)$$

M_f, it will be recalled is given by

$$M_f = \frac{f_d}{f_m} \qquad (3)$$

Substituting the value of M_f found in equation 3 into equation 2, there obtains

$$BW = 2 f_m \left(\frac{f_d}{f_m} + 1 \right) \qquad (4)$$

Performing the indicated multiplication, equation 4 becomes

$$BW = 2 (f_d + f_m) \qquad (5)$$

DOCID: 3726612

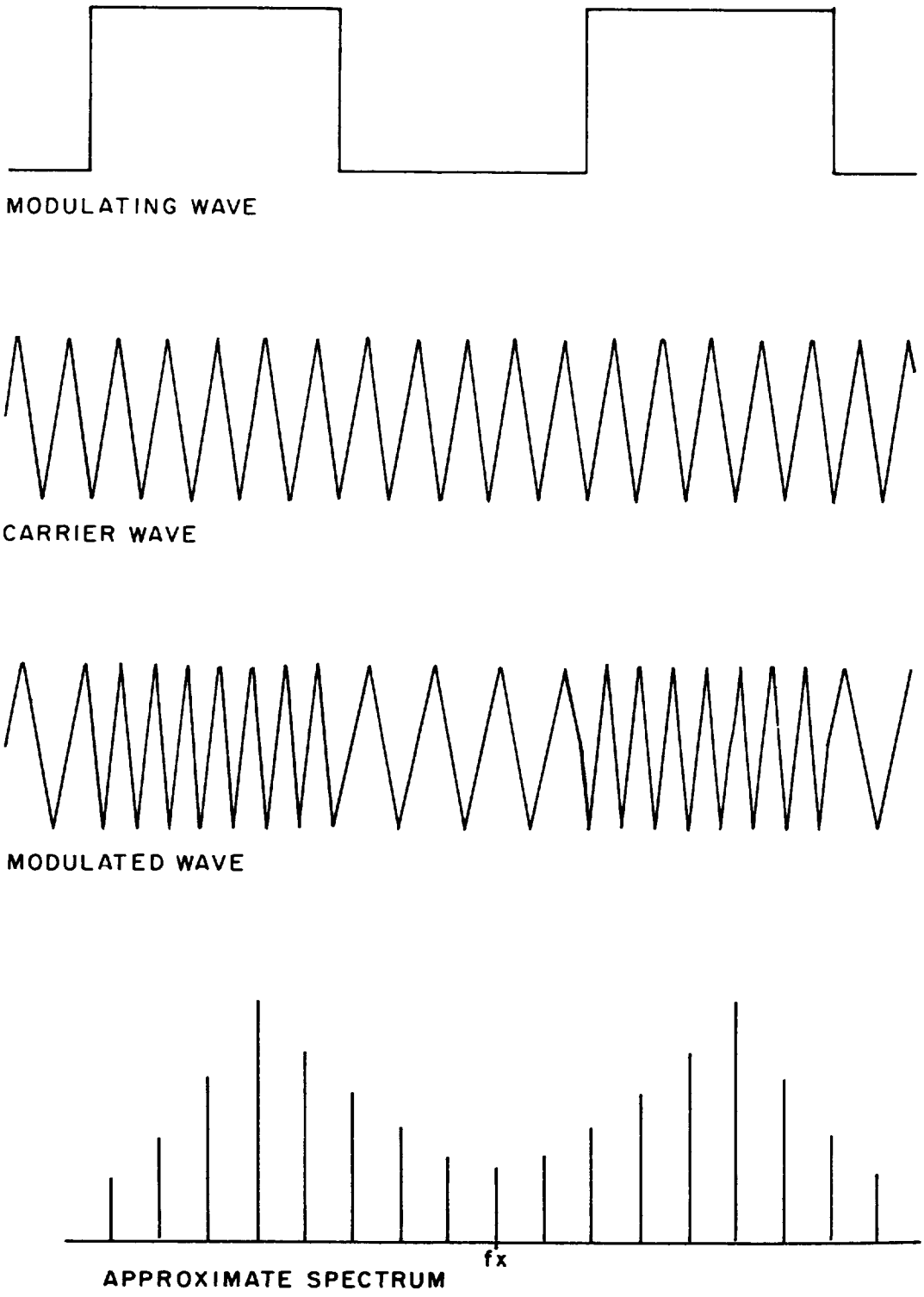


FIGURE 63. FREQUENCY MODULATION BY A SQUAREWAVE.

DOCID: 3726612

A. Wide Band F-M

Wide band F-M is the F-M encountered in commercial practice for public entertainment. Wide band F-M is distinguished by having modulation indices that are always much greater than unity. This form makes full use of the bandwidth equation, BW = $2(f_d + f_m)$. For example, for a frequency deviation ± 75 kc and a voice modulating wave whose upper frequency limit is 3.3 kc, the transmission bandwidth is,

$$BW = 2(f_d + f_m) = 2(75 + 3.3) = 156.6 \text{ Kilocycles}$$

B. Narrow Band F-M

For certain communication functions narrow band F-M is sometimes used. It is used extensively for police radio communication. Narrow band F-M is defined as having a modulation index of approximately unity. This type of F-M does not make full use of the F-M principle as evidenced by the bandwidth formula. For example, assume a system where $F_d = \pm$ 5 kc and the upper voice limit is 3.3 kc as before. (Notice here that the deviation is hardly more than the highest modulating frequency.)

$$BW = 2\ (f_d + f_m) = 2(5 + 3.3) = 16.6 \text{ Kilocycles}$$

This is only slightly more than twice the bandwidth required for an A-M signal carrying the same modulation.

C. Very Narrow Band F-M

This form of F-M makes almost no use of the F-M principle. The modulation index for this type of system is less than unity. A typical system for illustration is one where the deviation is 1 kc and the highest modulating frequency is again 3.3 kc.

$$BW = 2(f_d + f_m) = 2(1 + 3.3) = 8.6 \text{ kc}$$

The bandwidth required for this system compares very favorably with that of a similar A-M system.

VI. MATHEMATICAL POSSIBILITY OF P-M

No new arguments need to be presented to establish the possibility of phase modulation. Phase modulation is accomplished if the added phase angle, ϕ, in the sinewave equation, $a = A \sin 2\pi ft + \phi$, is acted upon by a modulating wave.

DOCID: 3726612

It is only necessary to contrive a means whereby the sinewave is distorted by varying the phase component of the angle θ in accordance with the amplitude changes of an information wave. By this process a complex modulated wave is produced and new frequencies are generated as sidebands.

VII. THE PHASE MODULATION WAVEFORM

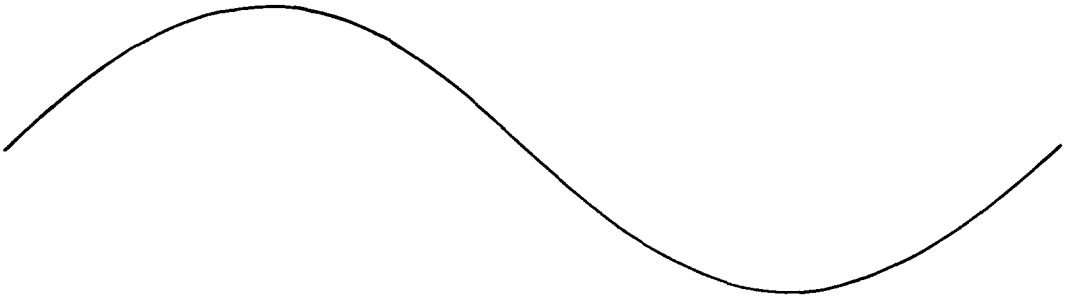
The typical phase modulated waveform produced by a sinusoidal modulating wave is illustrated in Figure 64. The waveform resulting from frequency modulation by the same wave is also shown in this figure for the sake of comparison. It is interesting to observe that the P-M wave looks like an F-M wave produced by a cosinusoidal instead of sinusoidal modulating wave. The P-M modulated waveform is that which would result if the rotating vector normally generating a sinewave were advanced in phase by some angle beyond its normally steadily increasing angle, and then retarded in phase by some angle to less than its normally steadily increasing angle. That is, it moves ahead of and then drops behind the steadily increasing angle that would exist in the absence of modulation.

A. Generation of the P-M Wave

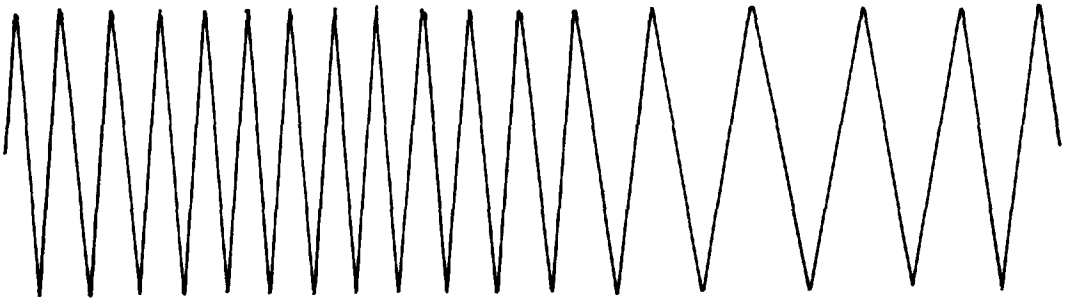
The phase changes produced in phase modulation are accomplished by the amplitude of the modulating wave. When the amplitude of the modulating wave is greatest, the phase deviation is greatest. When the phase is changing due to either an increasing or a decreasing modulating amplitude, there is some rate at which it changes. By definition, however, rate-of-change-of-phase is frequency; hence, when phase changes are taking place there is an accompanying frequency deviation. While the phase is advancing, the rate-of-change-of-phase is increased giving the appearance of many cycles occurring in a given interval of time. This results in high frequency during the change. While the phase is retarding, the rate-of-change-of-phase is decreased, giving the appearance of a few cycles stretched out over an equal interval of time. This results in low frequency during the change. These accompanying frequency deviations occur only while phase is changing.

Since the phase deviation is proportional to the variation in amplitude, the rate-of-change-of-phase is proportional to the rate-of-change-of-amplitude. The amplitude of a sinewave increases at a most rapid rate while passing through zero. The rate gradually decreases to zero as the sinewave reaches its peak. If the rates-of-change-of-amplitude were calculated and plotted for a number of points along a sinewave, the plots would indicate a cosinewave. In brief, the rate-of-change-of-amplitude of a sinewave is a cosinewave. For this reason, when modulating in P-M with a sinewave, the rate-of-change-of-phase (the accompanying frequency deviation) varies cosinusoidally. This is why the P-M wave

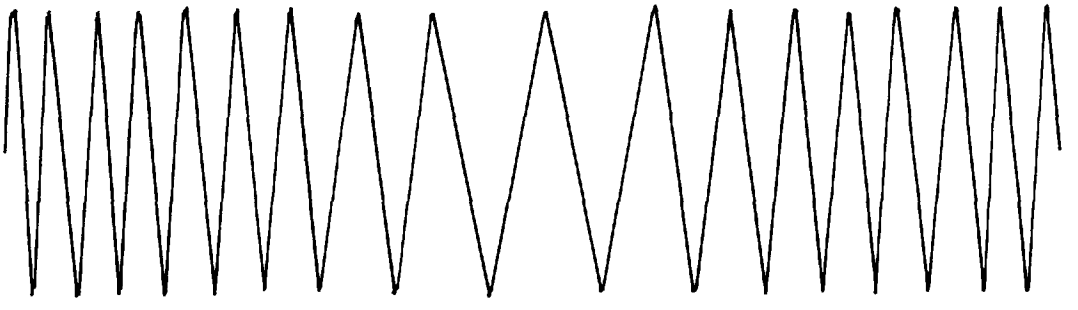
DOCID: 3726612



A. MODULATING WAVE.



B. FREQUENCY MODULATED WAVE.



C. PHASE MODULATED WAVE.

FIGURE 64. F–M & P–M WAVE COMPARED.

DOCID: 3726612

in Figure 64 looks like an F-M wave produced by a cosinusoidal instead of a sinusoidal modulating wave.

Figure 65 is similar to Figure 58 and illustrates that phase deviation in P-M bears the same relationship to modulating amplitude and frequency as frequency deviation in F-M bears to the modulating amplitude and frequency.

B. Analysis of the P-M Waveform

The similarity of the P-M and F-M waveform is no mere coincidence or accident. It must be remembered that both are forms of angle modulation which is defined as, "Modulation in which the angle of a sinewave carrier is the characteristic varied from its normal value," (53 IRE 11.S1). The similarity of F-M and P-M is therefore implied.

P-M is concerned with an angular distance (measured in radians) while F-M is concerned with an angular rate (measured in radians/second). Any change in angular motion must, of necessity, involve changes in both angular distance (phase) and angular rate (frequency). Therefore, any deviation in phase inherently involves a deviation in frequency and any deviation in frequency inherently involves a deviation in phase. Hence, phase modulation cannot occur without some accompanying frequency modulation and frequency modulation cannot occur without some accompanying phase modulation.

The feature that distinguishes P-M and F-M is whether phase deviation or frequency deviation is proportional to the modulating amplitude. In P-M the phase deviation is directly proportional to the modulating amplitude (and the accompanying frequency modulation is proportional to the modulating frequency) whereas, in F-M the frequency deviation is directly proportional to the modulating amplitude (and the accompanying P-M is inversely proportional to the modulating frequency).

The A-M modulated wave appeared to be a carrier whose amplitude was changing, but it was found that in reality it was a complex wave consisting of a carrier of constant amplitude, and side frequencies. The F-M modulated wave appeared to be a carrier whose frequency was changing but it was found that in reality it too was a complex wave consisting of a carrier (possibly) of constant frequency and of many sideband frequencies. It should not be surprising, then, to find that the P-M wave is also complex in nature and consists of a carrier of uniformly increasing phase and many sideband frequencies.

1. Sidebands in P-M

The same basic spectrum is generated for P-M as is generated

DOCID: 3726612

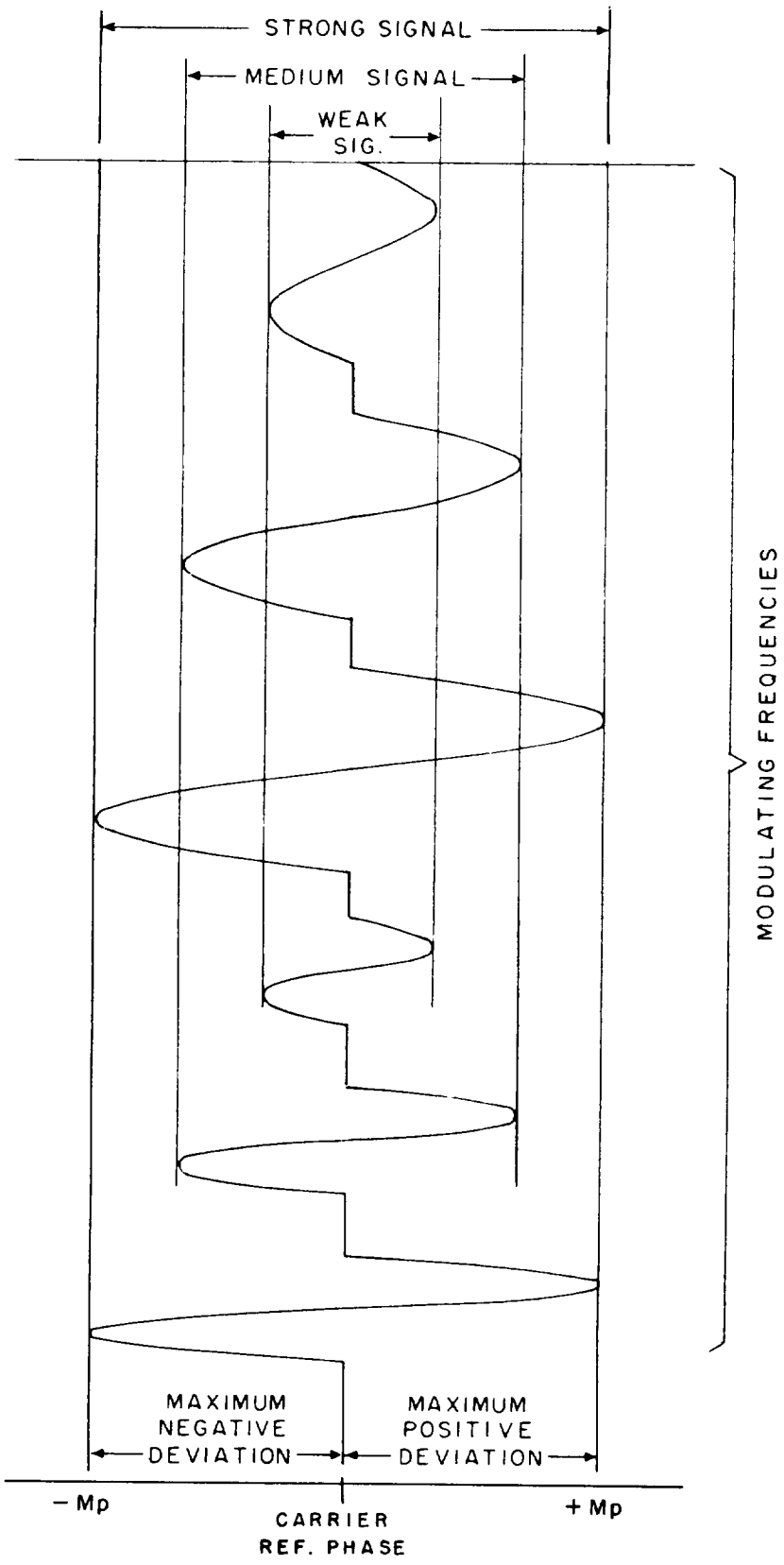


FIGURE 65. THE EFFECT OF MODULATING FREQUENCY & AMPLITUDE ON P–M.

DOCID: 3726612

for F-M. That is, unlike the A-M spectrum with its pair of side frequencies for each modulating frequency, there is a theoretically infinite number of side frequency pairs generated for each modulating frequency. As with F-M, the frequencies of the sideband components are equal to the carrier frequency plus and minus all harmonics of the modulating frequency. These are centered around the carrier frequency and are separated from the carrier and from each other by multiples of the modulating frequency.

2. Power in the P-M Wave

The power relationships in a P-M modulated wave are identical to F-M, namely that the power in a P-M wave is constant, power in the sidebands simply being extracted from the carrier.

3. Frequency Translation

Frequency translation in P-M is identical to F-M.

VIII. THE PHASE MODULATION SPECTRUM

A number of similarities have been noted in the two forms of angle modulation--more similarities, in fact, than differences. These similarities are to be expected since F-M and P-M are not basically different. In broadcasting functions, pure P-M is almost never used and in the United States no pure P-M stations for broadcasting functions are licensed at all. However, the so-called F-M stations that are licensed in the United States actually utilize F-M and P-M in a single hybrid system with frequency modulation being used for the part of the modulating band below one kilocycle, and phase modulation being used for the part of the modulating band above two kilocycles. The intermediate portion of the modulating band between one and two kilocycles is a transition range, and is transmitted by an angle modulation that is a compromise between pure F-M and pure P-M. The modulated spectrum is very similar for both and, for conditions to be described, the spectra are identical.

A. P-M and F-M Spectra Compared

Figure 66 illustrates the spectrum generated by a 15 kc sinewave phase modulating a carrier to produce a phase deviation (M_p) of 5 radians. It will be noticed that except for the labeling, this illustration is identical to Figure 60. This can and does happen because the spectrum is dependent upon the angular departure (phase deviation) that occurs in angle modulation. For phase modulation this angular departure or phase deviation is designated by M_p but phase deviation

DOCID: 3726612

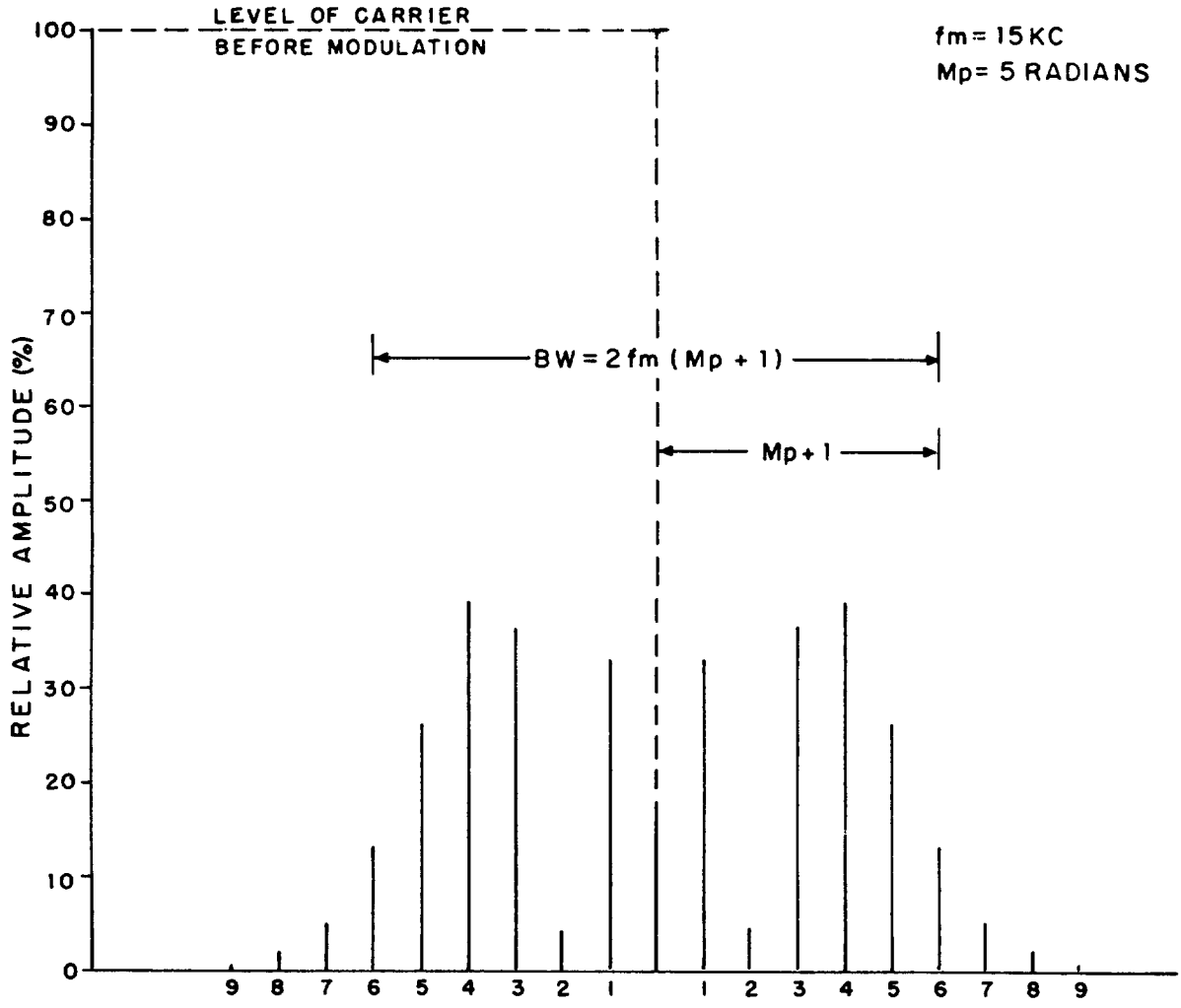


FIGURE 66. A PHASE MODULATION SPECTRUM.

DOCID: 3726612

for frequency modulation is only a secondary effect (the accompanying phase modulation) and is given by the modulation index, M_f. The M_p of phase modulation and the M_f of frequency modulation then are really the same thing--a measure in radians of the amount of phase deviation that is produced. In other words, a P-M wave that has a phase deviation of $M_p = 5$ (i.e., five radians) has the same angular departure as an F-M wave that has a modulation index of 5 and they both produce the same spectrum. Therefore, whenever the phase deviation (M_p) produced by a sinusoidal modulating wave in P-M is numerically equal to the modulation index (M_f) produced by a sinusoidal modulating wave in F-M, the spectra will be identical.

B. Phase Deviation and the Bessel Functions

Referring to Figure 61, it will be noted that the horizontal scale is labeled "Modulation Index (For F-M) or Phase Deviation (For P-M)." This is because the Bessel functions apply equally to P-M and F-M spectra. For P-M, it is only necessary to know M_p to use the Bessel function curves in the same manner that they were used for F-M.

IX. BANDWIDTH CALCULATIONS

Figure 66, like Figure 60, indicates that the spectral components having the greatest amplitude fall just inside the deviation limits. At least one more component (on each side) with considerable amplitude falls just outside the deviation limits. The number of spectral lines occurring to this point is equal to M_p (or M_f) + 1. Where M_p or M_f are equal to 5 this gives 6 spectral components on each side of the center frequency. The quantity $M_p + 1$ is the number of spectral components occurring in one-half the transmission bandwidth. The total bandwidth then must be twice this value or $2(M_p + 1)$. To interpret this in terms of frequency, $2(M_p + 1)$ must be multiplied by the modulating frequency, f_m. The P-M bandwidth equation then exactly parallels equation 2 presented for F-M bandwidth.

$BW \doteq 2f_m (M_p + 1)$ Phase Modulation

$BW = 2f_m (M_f + 1)$ Frequency Modulation

An example using the P-M bandwidth equation and the F-M bandwidth equation should make it clear that a P-M wave and an F-M wave, whose respective phase angle deviations (M_p and M_f) are equal to each other, have equivalent spectra and require the same transmission bandwidth for their spectra.

DOCID: 3726612

For the P-M wave

f_m = 5 kc

and

M_p = 2.5 (radians)

Bandwidth for the P-M wave is,

$BW = 2f_m (M_p + 1)$

$= 2x5 (2.5 + 1) = 35$ kc

For the F-M wave,

f_m = 5 kc

and

f_d = 12.5 kc

Computing the modulation index,

$$M_f = \frac{f_d}{f_m} = \frac{12.5 \text{ kc}}{5 \text{ kc}} = 2.5$$

(Notice that M_f is equal to M_p)

Bandwidth for the F-M wave is,

$BW = 2f_m (M_f + 1)$

$= 2x5 (2.5 + 1) = 35$ kc

The bandwidths are equal and the spectra are identical because the phase angle deviations of the two waves were identical.

Taking another example, there is a P-M wave where f_m = 15 kc and M_p = 5. The bandwidth when computed is 180 kc. Using the same highest modulating frequency, what frequency deviation (f_d) would an F-M wave have to have to produce an identical spectrum?

Solution:

M_f for the F-M wave must be equal to M_p if the spectra are to be identical. Therefore,

$$M_f = 5$$

It is known that in F-M, $M_f = \frac{f_d}{f_m}$, therefore

$$5 = \frac{f_d}{f_m}$$

It was said that f_m is 15 kc, therefore

$$5 = \frac{f_d}{15 \text{ kc}}$$

Transposing and solving for f_d,

$$f_d = 5 (15 \text{ kc}) = 75 \text{ kc}$$

DOCID: 3726612

Therefore, the f_d that an F-M wave modulated by the same modulating frequency would have to have in order to produce an identical spectrum is 75 kc. That the bandwidth is identical may be proven by use of either equation 2 or 5 for F-M bandwidth. Using equation 5,

$$BW = 2\ (f_d + f_m)$$
$$= 2\ (75 + 15) = 180 \text{ kc}$$

This is the same bandwidth that the P-M wave required.

CONCLUSION

Angle modulation, its two forms not basically different, provides a means of encoding information by generating a complex wave whose angle is varied from its normal value in proportion to the amplitude aspect of the modulating intelligence to be encoded. In keeping with this method of encoding, the rate at which the particular quantity (whether frequency or phase) is made to deviate is determined by the frequency aspect of the wave being encoded.

The resulting wave consists of a number of side frequencies, the aggregate of which would occupy a theoretically infinite bandwidth. In actual practice, a much smaller bandwidth is acceptable. The bandwidth equations submitted here are based on the assumption that sideband components whose amplitudes are less than 15% of the amplitude of the unmodulated carrier can be excluded without detrimental loss of information.

For frequency modulation and phase modulation bandwidth the calculations must be made to include the highest modulating frequency just as was required in amplitude modulation.

Along with A-M, the techniques of F-M and P-M find wide application in communications, guidance, and telemetry.

X. PRACTICAL EXERCISES

1. The transmission bandwidth of a wide band F-M wave is

   (a) double the intelligence spectrum
   (b) one-half the intelligence spectrum
   (c) $2\ f_m$
   (d) $2\ (f_d + f_m)$

DOCID: 3726612

2. The bandwidth of a narrow band F-M (NBFM) compares with

   (a) wide band F-M (WBFM)
   (b) P-M
   (c) A-M
   (d) nothing else

3. For a modulating wave having a given intelligence bandwidth, the narrowest transmission bandwidth is obtained with

   (a) narrow band F-M
   (b) amplitude modulation
   (c) wide band F-M
   (d) phase modulation

4. "Frequency swing" describes

   (a) the maximum total frequency excursion
   (b) the frequency deviation
   (c) the carrier frequency
   (d) the modulation index

5. The depth of modulation in F-M is given by

   (a) modulation index
   (b) highest modulating frequency
   (c) deviation ratio
   (d) per cent of maximum deviation used

6. In F-M the frequency deviation is

   (a) always accompanied by an amplitude variation
   (b) always accompanied by an amplitude variation and a phase deviation
   (c) never accompanied by either an amplitude variation or a phase deviation
   (d) always accompanied by a phase deviation.

7. In F-M, the frequency deviation is

   (a) the total excursion from the lowest to the highest frequency
   (b) one-half the total excursion from the lowest to the highest frequency
   (c) the modulating signal bandwidth
   (d) double the total excursion from the lowest frequency to the highest.

8. A phase deviation is

(a) always accompanied by an amplitude variation
(b) always accompanied by a frequency deviation
(c) never accompanied by a frequency deviation
(d) never accompanied by an amplitude variation

9. Normally a modulation index of approximately _____ indicates narrow band F-M

(a) 1
(b) 2
(c) 3
(d) 4

10. For F-M the wave at the center frequency

(a) varies in frequency
(b) varies in phase and amplitude
(c) varies in phase but not in amplitude
(d) does not vary at all

11. In an angle modulated wave

(a) all of the intelligence is in the carrier
(b) all of the intelligence is in the upper sideband only
(c) all of the intelligence is in the lower sideband only
(d) both sidebands contain <u>some</u> but neither sideband contains <u>all</u> of the intelligence

12. The same bandwidth is required for an F-M and a P-M wave having the same modulating frequency if

(a) $f_d = f_m$
(b) $f_d = M_p$
(c) $M_f = f_d$
(d) $M_f = M_p$

DOCID: 3726612

13. A modulating system that produces a complex modulated wave whose instantaneous peak amplitude is proportional to the instantaneous amplitude of the modulating wave is

(a) A-M
(b) P-M
(c) F-M
(d) there is no such system

14. In angle modulation, the energy in the sideband

(a) is constant
(b) comes from the modulating wave
(c) comes from the carrier wave
(d) depends upon the deviation ratio

15. A modulating system that produces a complex modulated wave whose phase deviation is proportional to the instantaneous amplitude of the modulating wave is

(a) A-M
(b) P-M
(c) F-M
(d) there is no such system

16. A certain P-M wave has a deviation in phase of 2.5 radians and the modulating frequency is 20 kc. What is the bandwidth required by the wave?

17. If the P-M system were to be replaced with an F-M system and if it were desired that the same modulating wave produce the same spectrum and require the same bandwidth, what frequency deviation must be produced by the modulating wave in the F-M system?

18. A 625 cycles per second modulating wave in a certain F-M system produced a 1.25 kc frequency deviation. What is the modulation index for this wave?

DOCID: 3726612

What is the phase deviation produced?

19. In terms of relationship to the modulating wave, what distinguishes P-M from F-M?

20. Given a modulation index of 2 and the Bessel function curves, what is the approximate amplitude of the first order sidebands relative to unmodulated carrier?

DOCID: 3726612

PART III

MULTIPLEXING

Quite often in pure communications, in electronic control (e. g., guidance), and in obtaining data from remote sources (telemetry), it is both extremely desirable and necessary to transmit and receive multiple channels of information over a common medium, all at the same time. The scheme for simultaneously transmitting multiple channels of information is called multiplexing. Two basic methods have been developed to accomplish this, and the term multiplexing applies equally to both. One method is called frequency division multiplexing, or multiplexing by frequency sharing. It makes use of the principle of frequency translation. The other of the two methods is called time division multiplexing, or multiplexing by time sharing. It makes use of pulse techniques.

The student will find that in the study of the two principal types of multiplexing (other methods are possible but not widely used) every modulation technique presented in previous chapters in this text will be used to the fullest extent. Frequency division multiplexing will be treated first. The succeeding section will discuss the basics of time division multiplexing.

CHAPTER 6
FREQUENCY DIVISION MULTIPLEXING

Some of the great advantages of modulation are to be found in frequency division multiplexing. By means of modulation techniques simultaneous signals can be transmitted at the same time without mutual interference even though they may have similar or identical frequency components. This is done quite simply by causing each such signal to modulate separate carriers having frequencies different from one another. By this means the simultaneous information-bearing signals are translated to different positions in the spectrum and thus kept separate. In the process of frequency division multiplexing it is demonstrated that the use of carriers and modulation need not necessarily be at radio frequencies and that, in fact, the only need for modulation of radio frequency carriers at all is to make possible the radiation of signals through space.

The student should not find frequency division multiplexing difficult or baffling. For, in reality, nothing new will be presented. Again, no new techniques will be introduced, but a little extra emphasis will be given to techniques that the observant student will readily recognize as having been taught in the lessons on modulation. This discussion will consider some of the aspects of the frequency division multiplex signal as it applies to both communications and telemetry. No attempt will be made here to explain any particular frequency multiplex system although

DOCID: 3726612

some actual standards will be quoted and used for illustration. The real purpose will be to show forth the underlying principles of such systems; it will then be left to the instructor and student to apply these principles in accordance with their own immediate requirement.

I. INFORMATION THAT IS FREQUENCY MULTIPLEXED

Multiplexing may be employed whenever there are multiple sources of information whose data must be transmitted simultaneously. To some of these, time division is an adequate system of multiplexing but for others, frequency division presents fewer restrictions. Communication functions that customarily use frequency division multiplexing are telephone and teletype. There are non-communicaion functions also that employ frequency division multiplexing among which are certain guidance systems and telemetry.

Since the sources of information for telephone and teletype are readily apparent, no elaboration will be made on them at the moment, but some explanation should be given to telemetry sources which are probably not as apparent to the average reader. The basic requirement of telemetry and, in fact, the root meaning of the term itself is, "to measure from a distance." The types of information to be processed in telemetry are varied in nature but in almost every instance the information is a type upon which some measurement can be made. To name but a few, flight attitudes as pitch, roll and yaw; altitude; heading; airspeed; and fuel flow; internal and external temperatures; vibration; stress; strain; pressure; light; radiation; blood pressure and body temperature are all among the types of information that at some time or another provide data for telemetry. Information sources that provide data such as this would include the vehicle itself, such as a pilotless aircraft, missile, rocket, space probe or satellite; environment; and some warm-blooded occupant of the vehicle.

The variations listed are not of themselves electrical in nature and so the device that picks up the variation must be capable of converting it into some electrical quantity with an analogous variation. The instruments that perform this operation are variously called end-instruments, function generators, pick-ups and transducers.

A. Classes of Information

All information, regardless of where it is used or what its sources (hence both communication and telemetry functions are included) can be grouped into two classes, and perhaps even a third if an admixture of the two basic ones is considered as a separate class. The two basic classes are discrete information and continuous information. When both the message and the signal bearing the

DOCID: 3726612

message are separate, individual and distinct symbols (for example, pulses), the information is said to be discrete. When both the message and the signal conveying the message are continuous functions of time (the term varying wave consisting of an infinite number of instantaneous values which may exist anywhere between the maximum and minimum amplitude limits), the information is said to be continuous. The third classification is termed mixed information and consists of both discrete amd continuous classes.

B. Rates of Information

The actual limits of the information rates that occur in multiplexed communication systems will vary depending upon whether the particular system is one of voice, teletype, entertainment communication, facsimile or some other. In telemetry it is possible to be more specific and give some limiting values.

The rates of telemetry data start with d-c and extend upward to as high as 10 kilocycles per second. This includes the harmonics of certain discrete data. Information whose rates extend over this range is called dynamic information. The information rates above 2100 cycles per second are, generally, not multiplexed because the higher information rates exceed the bounds set by the U. S. Telemetry Standards and to use a greater range while still observing the standard would greatly reduce the number of multiplex channels available.

When it is necessary to telemeter information at rates above 2100 cycles per second, it is usually necessary to provide an extra radio link, or in the instance of a recoverable vehicle, to incorporate a wide band recorder within the vehicle to record the higher rates. It is seldom necessary to use all of the wide dynamic range provided for in telemetry, however, because about 90% of the data producing sources characteristically yield information at rather slow rates. The slow rates belong, primarily, to the range of 0 to 10 cycles per second. Slowly varying values such as these are designated as quasi-static (almost stationary).

II. BASEBAND SYNTHESIS (FREQUENCY DIVISION MULTIPLEXING)

The information that is adaptable to frequency division multiplexing may be continuous, discrete, or mixed; it may range over the quasi-static through the dynamic rates; and it may represent some form of pure communication, guidance, or telemetry. Whatever the information source, the data for each multiplex channel is simply made to modulate a carrier associated with that particular channel. The modulation method employed may be A-M, A-M SSB, FM, or PM and may include any of the variations that exist for these. In practice some restrictions are made for sake of economy of spectrum, size and weight, expense, and efficiency.

DOCID: 3726612

Frequency division is a method of separating individual channels by associating each channel with some distinct frequency. In a word, it makes use of the principle of frequency translation as accomplished by modulation techniques. The result in any instance is a complex modulation system. The aggregate of all the translated channels is a composite signal, indeed a complex wave, whose spectrum bandwidth extends to the highest significant sideband of the channel having the highest carrier frequency. The band of frequencies occupied by these channels, regardless of its width or the number of channels involved, is called a baseband. Frequency division multiplexing is the synthesis of the baseband.

A. The Baseband in Communication

The function of a baseband in a communication system is to multiplex a number of channels of intelligence by causing the several channels to share the frequency spectrum. The baseband consists of carrier frequency or center frequency assignments and the various sideband spectra that develop around them due to modulation. The carriers occupying the baseband are usually called subordinate carriers or subcarriers because they are subordinate in both frequency and application to the final R-F carrier which is required if the system's signals are to radiate.

The baseband may include any type of modulation in any of the channels that compose it. The bandwidth occupied by a baseband depends upon the system in use and may extend from the lower audio range (even down to and including d-c as in facsimile systems) up into the megacycle region.

1. A Representative Frequency Multiplex Communication System

Figure 67 is a block diagram of a representative but entirely fictitious communication system illustrating the generation and use of a baseband. Suppose that there are available several pure audio tones whose frequencies are assumed to be 750, 1500, 2250, and 3000 cycles per second. These frequencies are illustrated in Figure 68A as a spectrum presentation. It will be noted that one of these tones is associated with each of the channels in Figure 67. These will serve either as carrier or center frequencies for their respective channels and each will be modulated by a separate source of information. To distinguish the channels, they shall be referred to as channel 1, channel 2, etc.

Suppose that it is desirable to transmit a teletype signal consisting of 4.3 millisecond pulses. The fundamental frequency of the equivalent squarewave for this signal is given by the reciprocal of twice the pulse duration and is approximately 116 cycles per

DOCID: 3726612

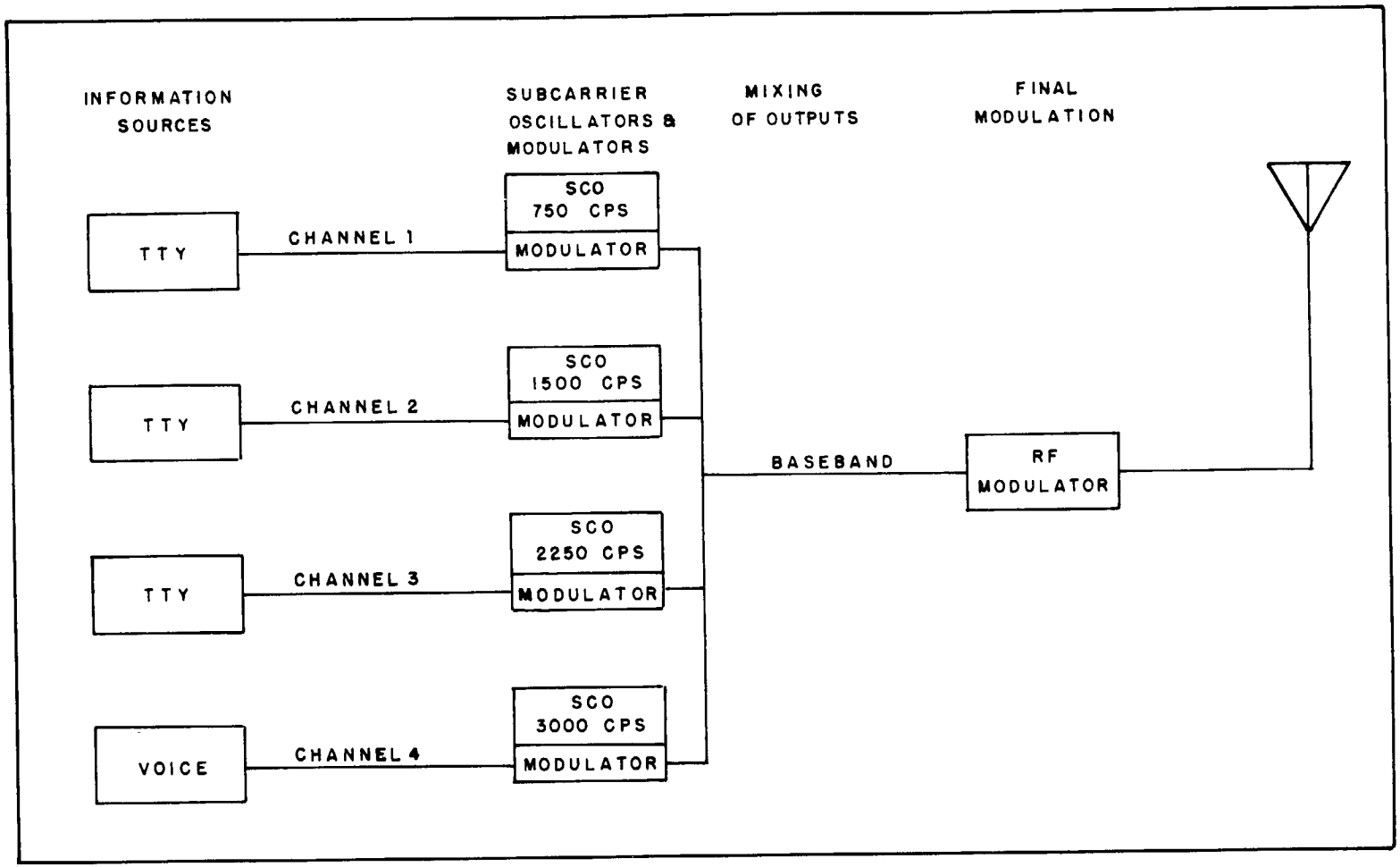


FIGURE 67. A REPRESENTATIVE FREQUENCY MULTIPLEXED COMMUNICATION SYSTEM.

DOCID: 3726612

second. This frequency and the third and fifth harmonics of it are indicated in Figure 68B.

The teletype signal will be used to frequency shift key channel 1 (the 750 cps center frequency). The mark condition will cause the frequency to deviate 85 cps above the channel center frequency, and the space will cause the frequency to deviate 85 cps below the center frequency, for a total frequency shift of 170 cycles per second. The squarewave modulation index may be found by the ratio of the frequency shift to the transmission speed in bauds, or by the product of the frequency shift and the baud length (in seconds). (It is also equal to the deviation divided by the PRF.) Any of these methods will give the same numerical result $SM_f = .73$. The resultant spectrum (an approximation) is shown in Figure 68C.

If the first three channels bear similar information (each from a different source, of course) the composite spectrum after modulation will closely resemble that shown in Figure 69A.

Channel four has a subcarrier whose frequency is three kilocycles. To add variety to the example under development, and to illustrate that a baseband can contain more than one type of modulation and more than one type of information, assume that this channel is to be amplitude modulated by voice intelligence with an effective bandwidth of three kilocycles. An approximation of the voice spectrum is shown in Figure 69B.

By reason of the three kilocycle carrier and amplitude modulation, the voice spectrum is translated in frequency resulting in an upper sideband with a normal spectrum (like that of the modulating wave) but ranging above the carrier, and a lower sideband occupying the same band as the modulating spectrum but with the frequencies inverted. By suppressing the lower sideband and the subcarrier a single sideband is left that is an exact reproduction of the voice spectrum but having each spectral component occur three kilocycles higher in frequency than it did before modulation.

2. Versatility of the Baseband

The model baseband is now complete. The four channels with their respective sideband spectra are shown in Figure 69C. The three teletype channels utilize FSK which is a form of F-M and the voice channel utilizes A-M single sideband with the carrier

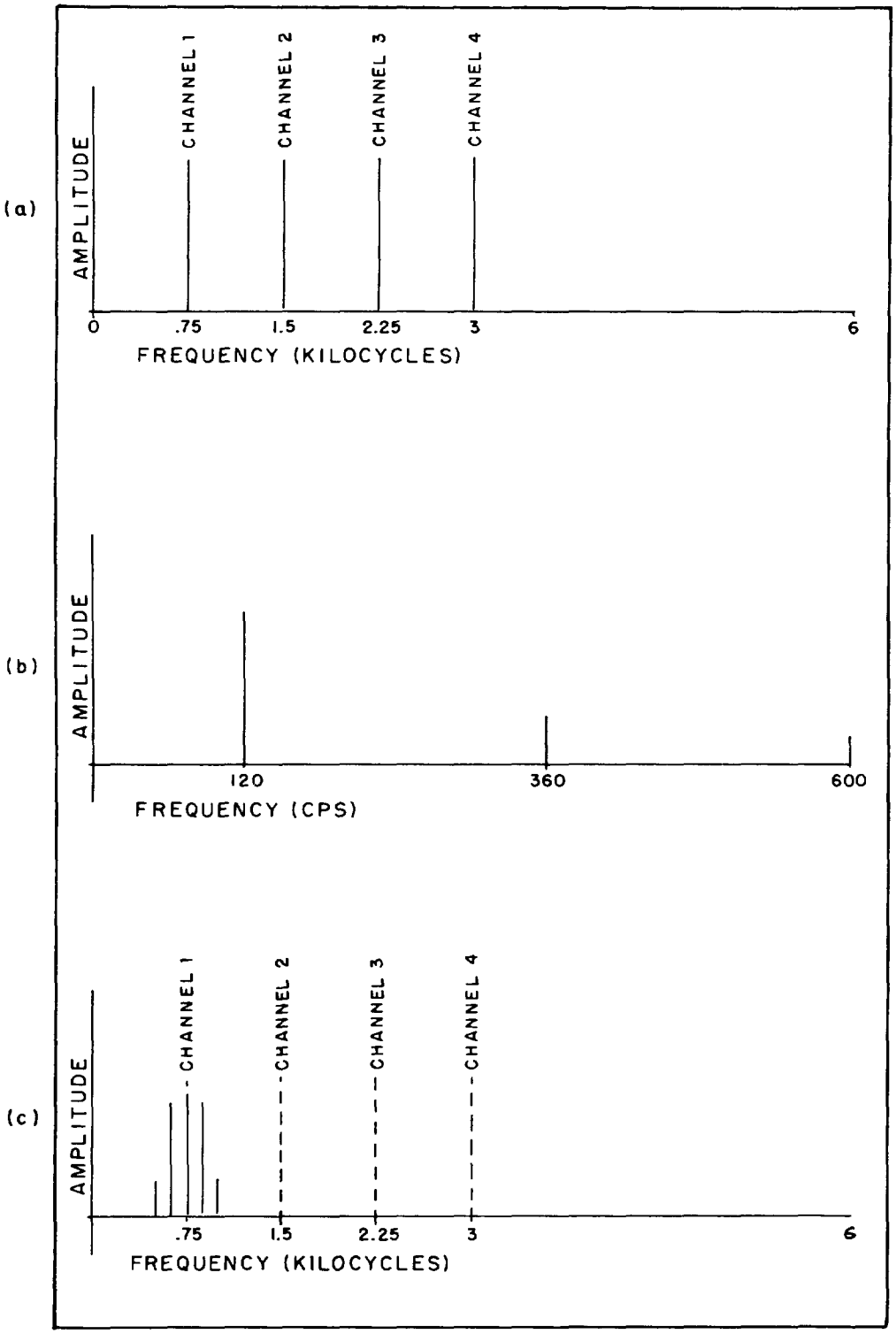


FIGURE 68. MODULATING, AND BASEBAND SPECTRA.

DOCID: 3726612

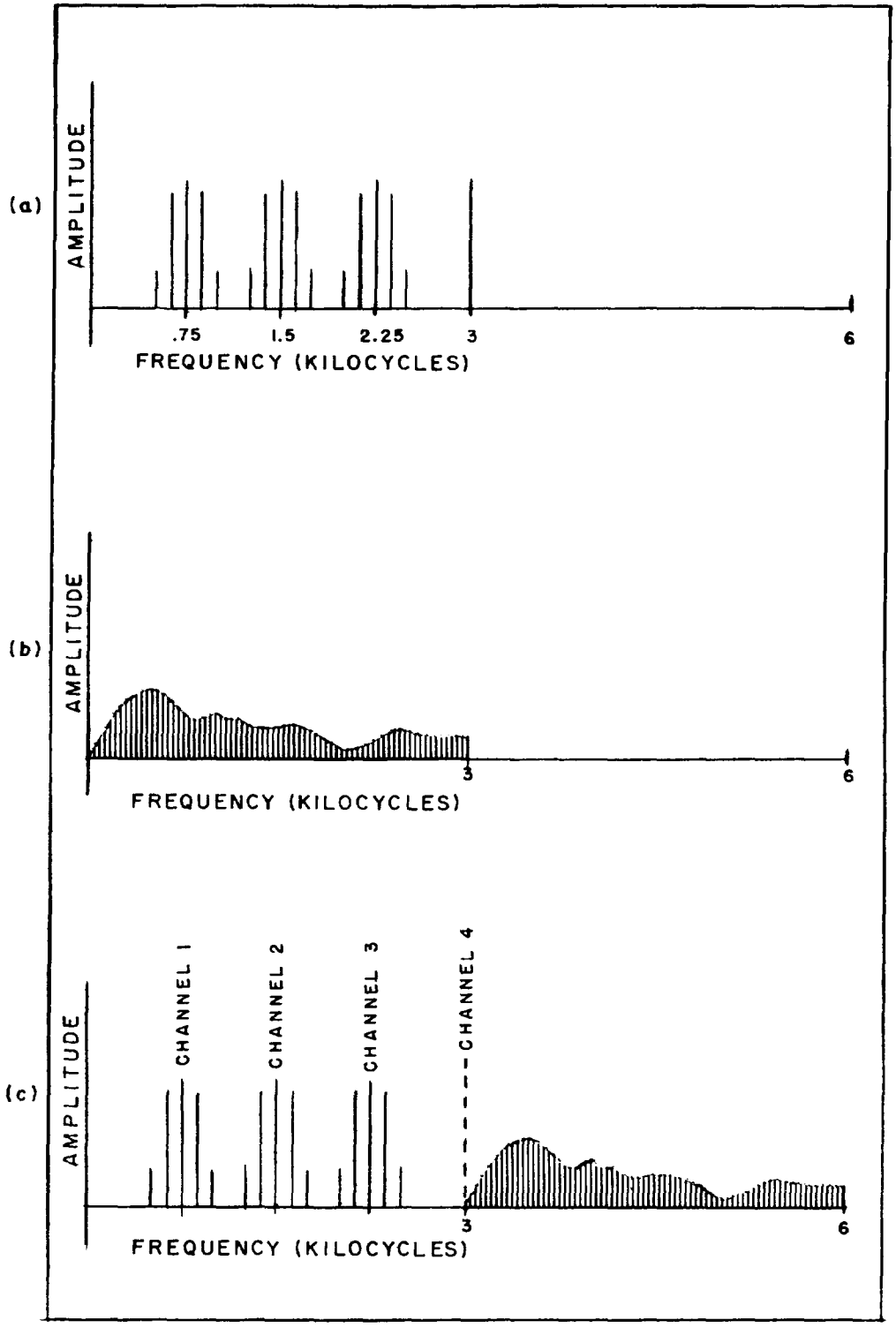


FIGURE 69. MODULATING, AND BASEBAND SPECTRA.

DOCID: 3726612

suppressed. This particular baseband, then, has two different types of information (three channels of discrete and one channel of continuous information) and two different kinds of modulation (both F-M and A-M). Thus it has been demonstrated that there is no restriction as to the type of information or kind of modulation employed in the baseband.

Typically the information applied to a baseband may be voice, teletype, telemetry and facsimile and may employ one or more kinds of modulation with their adaptations such as single sideband, on-off keying, narrow band F-M, P-M, or frequency shift keying. It will be noticed that the only thing that separates the various channels of intelligence after modulation is the frequency of the subcarrier associated with each channel. The subcarriers are spaced in frequency in such a manner that there is no overlapping of the sidebands produced in the modulation process. Thus, the sharing of the frequency range over a given band accomplishes frequency division.

B. The Baseband in U. S. Telemetry

In U. S. Telemetry, a standard has been established that limits the type of subcarrier modulation in frequency division multiplexing to F-M. This same standard also limits the number of subcarriers that may be used in the baseband, defines the maximum frequency deviation for each subcarrier channel, and fixes the center frequencies for the subcarriers. The tabulation of these limitations as they appear in the standard is reproduced on page 184.

DOCID: 3726612

TABLE I. SUBCARRIER BANDS

Band	Center Frequency (cps)	Lower Limit (cps)	Upper Limit (cps)	Maximum Deviation (percent)	Frequency Response* (cps)
1	400	370	430	7.5	6.0
2	560	518	602	7.5	8.4
3	730	675	785	7.5	11
4	960	888	1,032	7.5	14
5	1,300	1,202	1,398	7.5	20
6	1,700	1,572	1,828	7.5	25
7	2,300	2,127	2,473	7.5	35
8	3,000	2,775	3,225	7.5	45
9	3,900	3,607	4,193	7.5	59
10	5,400	4,995	5,805	7.5	81
11	7,350	6,799	7,901	7.5	110
12	10,500	9,712	11,288	7.5	160
13	14,500	13,412	15,588	7.5	220
14	22,000	20,350	23,650	7.5	330
15	30,000	27,750	32,250	7.5	450
16	40,000	37,000	43,000	7.5	600
17	52,500	48,560	56,440	7.5	790
18	70,000	64,750	75,250	7.5	1,050
A**	22,000	18,700	25,300	15	660
B	30,000	25,500	34,500	15	900
C	40,000	34,000	46,000	15	1,200
D	52,500	44,620	60,380	15	1,600
E	70,000	59,500	80,500	15	2,100

*Frequency response given is based on maximum deviation and a deviation ratio of five (see text for discussion).

**Bands A through E are optional and may be used by omitting adjacent bands as follows

Band Used	Omit Bands
A	15 and B
B	14, 16, A and C
C	15, 17, B and D
D	16, 18, C and E
E	17 and D

It is seen from Table I that the telemetry baseband consists of eighteen subcarrier assignments with center frequencies ranging from .4 kilocycles to 70 kilocycles. For ease of identification the subcarriers are numbered from one through eighteen in the order of increasing frequency.

DOCID: 3726612

Table I indicates a frequency response (i.e., highest modulating frequency, f_m) for each subcarrier band and also indicates a maximum frequency deviation (i.e., f_D) for each. This frequency deviation is different for each subcarrier channel and amounts to ± 7.5% of the subcarrier frequency for bands one through eighteen. For example, band 1 has a center frequency of 400 cycles per second. Plus and minus seven and one-half percent of this gives a frequency deviation of ± 30 cps. Subtracting the deviation from the subcarrier center frequency (400 - 30 = 370) gives the lower frequency limit, and adding the deviation to the center frequency (400 + 30 = 430) gives the upper frequency limit of this subcarrier band. This explanation has included all the quantities listed for subcarrier band 1 as they appear in Table I.

The values that appear in Table I have been selected so that the deviation ratio of the subcarrier modulators can be set at 5. (It will be recalled that the deviation ratio is a system characteristic used to establish the maximum circuit bandwidth.) This may be proven easily for any particular band by solving for the frequency deviation (column 2 times column 5) and dividing by the frequency response (highest modulating frequency found in column 6). When full use is made of this deviation ratio, even the highest modulating frequency within the frequency response of the channel (See Table I) can (depending upon the modulation amplitude) produce a modulation index of 5. In this manner provision is made for all subcarrier bands to carry wideband frequency modulation (WBFM)

1. Width of the Telemetry Baseband

The highest modulating frequency for the lowest frequency subcarrier is 6 cycles and the maximum deviation for this subcarrier is 30 cycles per second. The bandwidth, when determined by the formula taught in Angle Modulation is,

$$BW = 2(f_d + f_m) = 2(30 + 6) = 72 \text{ cps.}$$

The lower-most frequency in the baseband, then, is the center frequency of band 1 (namely, 400 cps) diminished by one-half the bandwidth of the subcarrier band, thus,

$$400 - \frac{72}{2} = 364 \text{ cycles per second.}$$

It will be noted that this is six cycles per second outside the lower deviation limit indicated in column 3 of Table I.

The highest modulating frequency for the numbered subcarrier having the highest frequency is 1.05 kilocycles and the maximum

DOCID: 3726612

deviation for this subcarrier is 5.25 kilocycles. The bandwidth determined by formula is,

$$BW = 2(f_d + f_m) = 2(5.25 + 1.05) = 12.6 \text{ kc}$$

The upper-most frequency in the baseband, then, is the center frequency of band 18 (namely, 70 kilocycles/sec.) plus one-half the bandwidth of this subcarrier band. This is,

$$70 + \frac{12.6}{2} = 76.3 \text{ kilocycles per second.}$$

It will be noted that this is 1.05 kilocycles per second outside the upper deviation limit indicated in column 4 of Table I.

The maximum bandwidth for the baseband determined by the requirements of subcarrier bands 1 through 18 can be as great as the spectral range between the lower and upper limiting frequencies just computed. This is 76.3 kc - .364 kc = 75.936 kc, or approximately 76 kilocycles/sec. It should be called to the readers attention that in normal practice not more than twelve of the eighteen subcarrier bands will be used at any one time because of the possibility of the sidebands of adjacent channels overlapping. Hence, the actual bandwidth of the baseband in any particular instance depends upon which subcarriers are used and upon the highest modulating frequency employed.

2. A Representative Frequency Multiplex System

In Figure 70 there is illustrated a representative telemetry system employing frequency division multiplexing. Here may be seen four information sources representing four distinct channels of information, the end-instruments, and the subcarrier oscillators with their modulators. The subcarrier oscillators each have a different frequency as required for frequency translation and stacking of channels so as to avoid overlapping of sidebands. The outputs are seen to combine to form the baseband which becomes the modulating wave for the final R-F carrier.

An interesting point to note is that for this representative system provision has been made for only three subcarrier bands. This means that there are more information channels than there are channels within the system. Channel one in this fictional system will be given to the measurement of very slowly varying

DOCID: 3726612

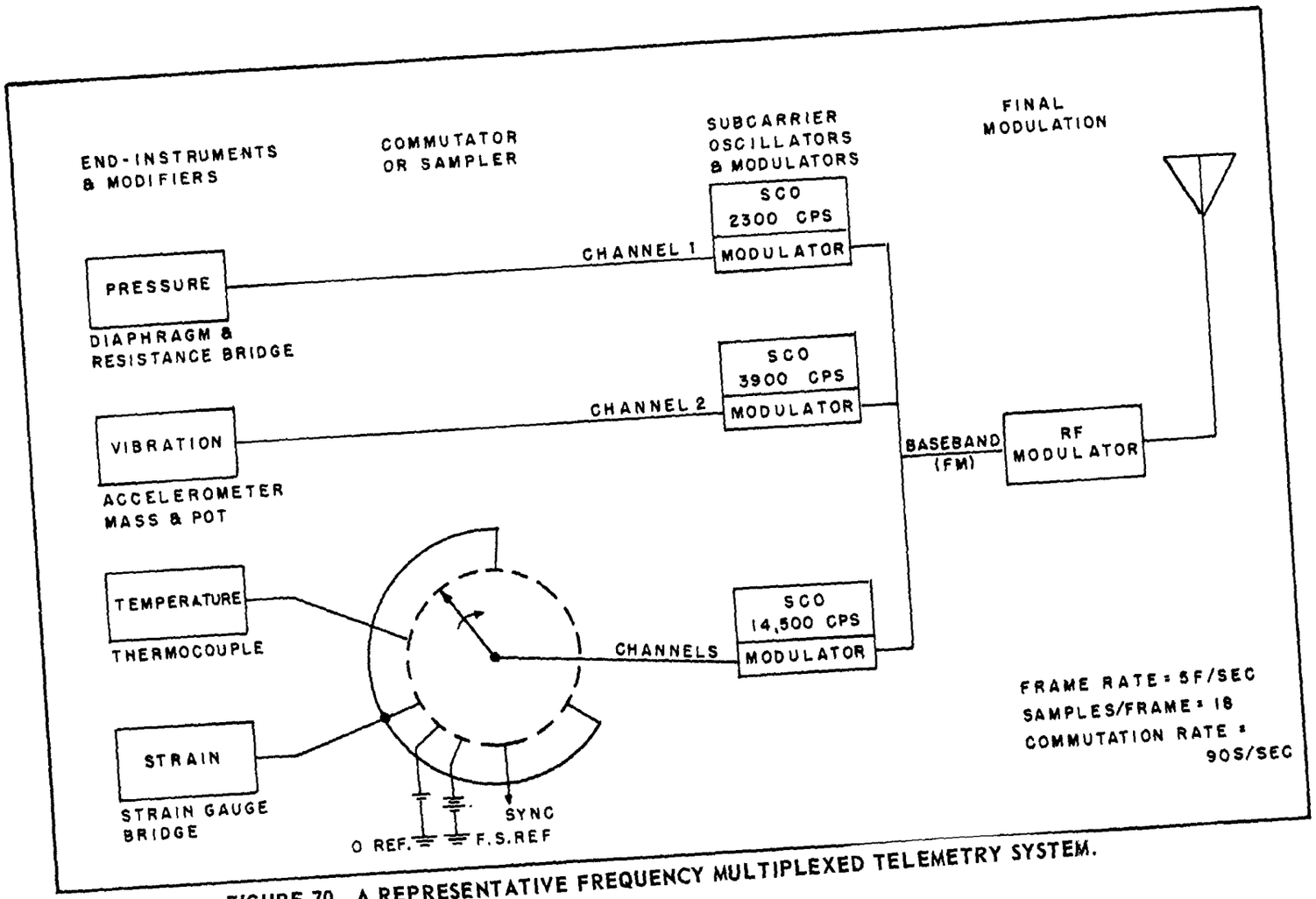


FIGURE 70. A REPRESENTATIVE FREQUENCY MULTIPLEXED TELEMETRY SYSTEM.

DOCID: 3726612

pressure changes. The changes will be assumed to be so slow as to be considered quasi-static. Since some pressure will always be present, it is continuous data. Following the block representing the information source, there is the end-instrument, the function of which is to convert the changes in pressure into an electrical analog (i. e., an electrical signal that varies at the same rate, and by the same percentage as the change in pressure). By means of a modulator, this electrical analog is used to frequency modulate the output of a subcarrier oscillator. Band 7 (Refer to Table I) was selected for this purpose. This subcarrier and the sidebands resulting from modulation by the information become a part of the baseband.

Channel two is similarly constructed. The information source in this instance produces vibration. Vibration will be considered to be continuous, to be variable over a fairly large frequency range, and to have some transient amplitude characteristics. This, then, is continuous but dynamic information. The end-instrument appropriate to this variation might be an accelerometer. The output of the accelerometer is an analog variation used to modulate a subcarrier in the same manner as with the previous channel. Band 9 (Refer to Table I) was selected here. This, too, becomes part of the system baseband.

So far, only pure frequency division multiplexing has been mentioned. In many practical instances, however, one of the schemes of time division multiplexing is employed in conjunction with frequency division. The object of the double multiplexing is to increase the channel capacity. The scheme employed is called sampling. Sampling will be discussed in more detail in the subsequent lesson on time division multiplexing, but a brief explanation must suffice at present. Sampling is a means of reducing a continuous information wave to pulses which are samples of the amplitude of the wave, and which, by their very nature, are transmitted intermittently (but at regular intervals) rather than continuously. The specification of the information can be carried in these discrete pulses. Either mechanized or electrical means may be employed to produce the samples, but both methods of sampling are generally referred to by the term commutation.

Sampling is the process used to allow the two remaining information channels in Figure 70 to share the one remaining channel of the system. The time sharing samples that result are then used

DOCID: 3726612

to modulate the subcarrier just as any other channel would and thus the information is transferred to the baseband. Since these two information channels do not individually compose a channel within the system, they may be classed as subchannels and the commutation or sampling process used to accomplish this is called sub-commutation. Each pole on the commutator represents a distinct subchannel. Eighteen such poles exist in this system but four of them are used for synchronization and calibration purposes.

The third information source produces a very slowly changing temperature. The information is continuous and definitely quasi-static. A thermocouple would be a suitable end-instrument to convert these changes into an electrical signal. The analog voltage from the thermocouple, instead of being fed directly to the modulator for channel three, is fed to one pole of the sampler. The commutator pole corresponding to subchannel four has been arbitrarily chosen for this input signal. The signal is sampled once for every revolution of the commutator rotor.

The fourth end-instrument picks up information produced by mechanical strain. This is continuous quasi-static information (in this instance) and is easily picked up by means of a strain gauge bridge. Like the electrical signal produced by the last previous end-instrument, it is not fed directly to the subcarrier modulator, but is fed to the commutator where it is sampled. There is now a decided difference: this signal is not fed to just one commutator pole but to three different poles that have been cross-connected (strapping, as it is called) so that the signal will be sampled three times for every one revolution of the commutator rotor. The three subchannels occupied by this information channel are one, seven, and thirteen. It should be noticed that the three poles are equally distant from each other.

All the data applied to the poles of the commutator, whether synchronization information, calibration information, or data to be measured, is sampled each time the rotor makes its revolution. The output which is taken from the rotor, is a time multiplexed signal which becomes a single input to the subcarrier modulator. By virtue of the sampling process, this input consists of discrete data in contrast to the inputs to the other subcarrier modulator which were continuous. Modulation of the subcarrier channel is no different than before. Band 13 (Refer to Table I) was selected here.

DOCID: 3726612

The output of the subcarrier modulator is added to the baseband of the multiplex system.

With this last addition, the baseband of the representative frequency multiplex system is complete, and the baseband to be considered as a complex modulating wave is ready to either frequency modulate or phase modulate the final R-F carrier so that the information can be radiated through space to suitable pickup stations where the data can be demultiplexed and analyzed into measurements.

3. Why Double Multiplexing is Used

The student may wonder why a process of double multiplexing was used in the representative system just analyzed. This concept was included here because it is frequently done in actual systems. Table II, reproduced from the Telemetry Standards for Guided Missiles, deals with the requirement for time multiplex used in a frequency multiplex system such as was used in the example here. The table is reproduced on page 191.

DOCID: 3726612

TABLE II. COMMUTATION SPECIFICATIONS FOR AUTOMATIC DECOMMUTATION

No. Samples Per Frame*	Frames Per Second	Commutation Rate** Samples Per Second	Lowest Recommended Subcarrier Bands (cps)
18	5	90	14,500
18	10	180	22,000 (±15%) or 30,000 (±7.5%)
18	25	450	30,000 (±15%) or 70,000 (±7.5%)
30	2.5	75	10,500
30	5	150	22,000 (±7.5%)
30	10	300	22,000 (±15%) or 40,000 (±7.5%)
30	20	600	40,000 (±15%)
30	30	900	70,000 (±15%)

*The number of samples per frame available to carry information is two less than the number indicated, because the equivalent of two samples is used in generating the frame synchronizing pulse.

**Frame rate times number of samples per frame.

The eighteen sample frame, ninety sample/second commutation system appearing in the first row of the table is that which was used in the example.

The reason for time multiplexing information in a frequency multiplex system is to group the information from several sources onto a single subcarrier band. The method can be used to greatly reduce the number of channels needed in the multiplex system to handle the information from some given number of sources. The disadvantage in this form of multiplexing is that any subcarrier channel employing subcommutation is limited to quasi-static data from sources producing continuous information.

When subcommutation is applied to the input derived from several information sources, the data from one or more of the sources may vary at a sufficiently high rate that it cannot be properly recovered from samples taken at the slow sampling rate of the commutator. This can be corrected by taking more

DOCID: 3726612

than one sample for each commutator revolution. This technique is called strapping and is the method employed for the fourth information source in the example of Figure 70. Strapping by cross-connecting commutator poles has the advantage of increasing the sampling rate but has the disadvantage of decreasing the number of sources that can be sampled because it reduces the number of available commutator poles.

III. THE BASEBAND AS A MODULATING WAVE

For both a communication system and a telemetry system which use frequency multiplex methods, a discussion has been given which describes how the various channels of information may be multiplexed by frequency translation and frequency division. Not all possible waves, inputs, or spectra have been mentioned or illustrated, nor should it be necessary to do so because the same basic principles that have been taught underlie all methods regardless of their source or purpose.

A. The R-F Carrier

The R-F carrier used in any system of multiplexing has no peculiar qualities that make it different or distinct from an R-F carrier used for other electronic functions where radiation through space is required. Depending upon the application at hand, a multiplexed signal may not even require a final R-F carrier. For example, some functions do not need to be radiated but the signal can be carried by wire from its source to its destination, or in other instances the baseband can be recorded by a video recorder. If, on the other hand, radio transmission is required, an R-F carrier that is sufficiently high in frequency to radiate effectively and final modulation are necessary.

B. Final Modulation

Whether a frequency multiplex system is used for communication or telemetry, whether the subcarriers are modulated A-M, F-M or some variation of these, or whether the system is purely frequency multiplex or if subcommutation is used prior to modulating a subcarrier, the final result is a baseband of subordinate carriers and their sidebands. This composite signal (the baseband spectrum) may be thought of quite simply as just another modulating wave which must be applied to the R-F carrier.

The kind of modulation employed for the final step before transmission could be any of the existing forms or their modifications, all of which are used in various communication functions. However, a fact of which the student should already be aware is that the different modulation schemes have their own bandwidth characteristics which may impose advantages or disadvantages on their

DOCID: 3726612

usage. In the instance of the model communication system of Fig. 67 and whose baseband spectrum appears in Fig. 69, three approximate R-F spectra are shown in Fig. 71 to illustrate the results when the baseband is used to accomplish amplitude modulation, single sideband suppressed carrier, and frequency modulation of the R-F carrier.

In the United States, the established practice and standard for guided missile telemetry is to use angle modulation, usually F-M, for modulating the R-F carrier. Several bands have been designated for the final carriers. These are 216 to 260 megacycles per second; 1435 to 1535 megacycles per second with half the band relegated to aeronautical flight testing and the other half of the band relegated to missile and space vehicle testing; and 2200 to 2300 megacycles per second. Including drifts and all other variations the carriers operating within these bands must maintain a stability of .01% or better. The separation between carriers depends upon the band, or part of the band, that the carrier is in but the bandwidth permitted for the channel around any particular carrier is 500 kilocycles per second.

The maximum frequency deviation permitted in the case of F-M (or equivalent frequency deviation in the case of P-M) is ± 125 kilocycles per second, and the recommended minimum deviation is ± 75 kilocycles per second. When the maximum frequency deviation is the one used, the guard band occurring at each end of the telemetry channel is 125 kilocycles per second.

Although provision is made for the subcarriers in telemetry to be modulated as wide band frequency modulation (WBFM) no such provision is made for the modulation of the final R-F carrier. The deviation ratio of the R-F modulator is set at about 1.8 . This ratio is based upon the maximum allowable deviation (± 125 kc) and the highest system subcarrier frequency (rather than upon the actual highest baseband frequency). Thus,

$$D = \frac{f_D}{f_{sco}} = \frac{125 \text{ kc}}{70 \text{ kc}} \approx 1.8$$

This low value for the deviation ratio is necessary in order to limit the radio frequency bandwidth.

The effect that this equipment deviation ratio has is that the highest available subcarrier and its sidebands produce narrow band F-M in the final modulation, whereas the lower subcarriers and their sidebands produce varying degrees of wide band F-M.

DOCID: 3726612

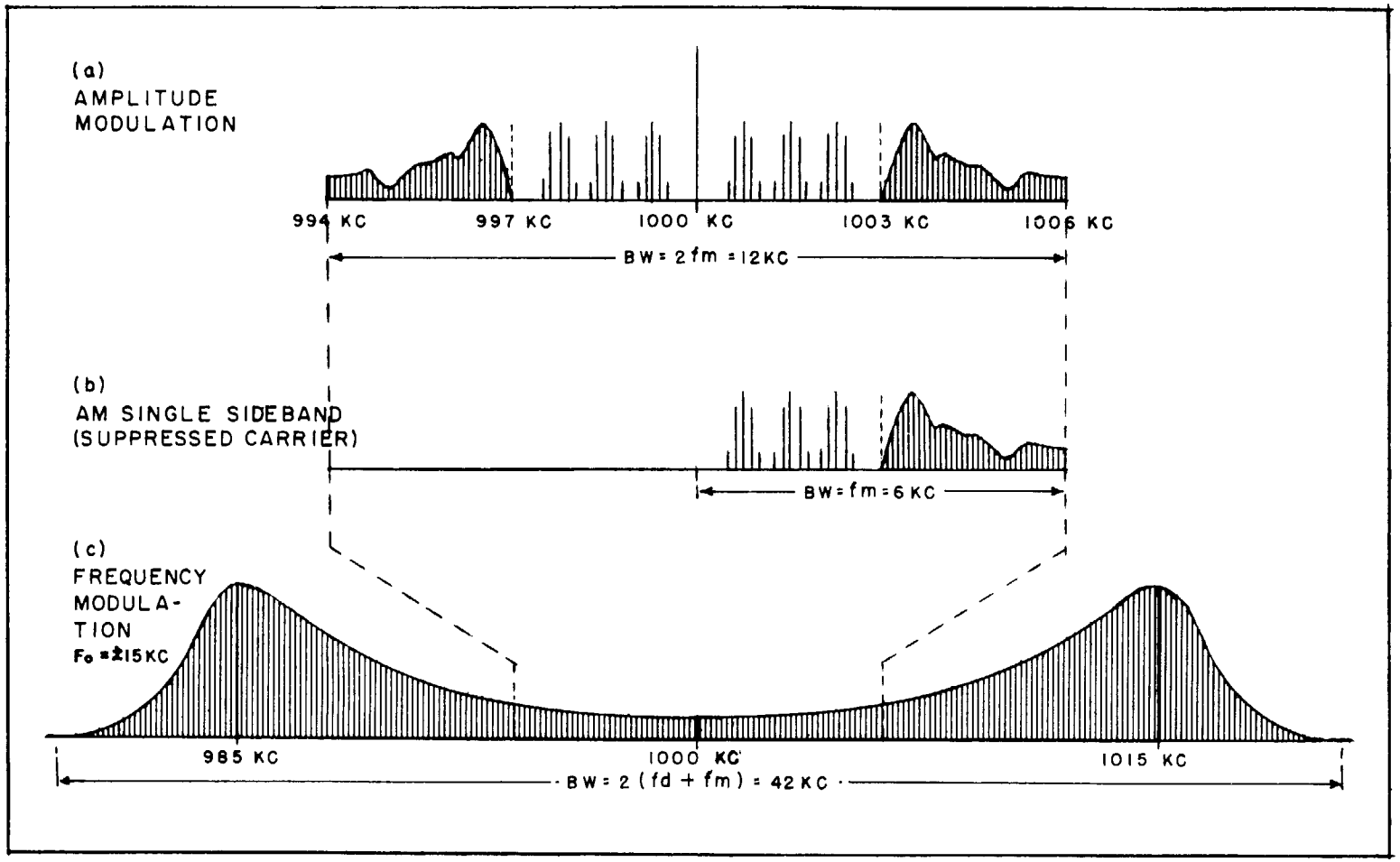


FIGURE 71. FINAL MODULATION OF THE BASEBAND.

DOCID: 3726612

C. Multiplex System Designations

It is customary to designate multiplex systems by the letter-groups that identify the forms of modulations used. These letter-groups are the conventional ones: A-M for amplitude modulation, P-M for phase modulation, and so on. A letter-group appears in the multiplex designation for each form of modulation used. The order in which these letter-groups appear is indicative of the order in which the modulation takes place, starting with the modulation produced by the individual input channels and working toward the modulation used for the final carrier. Looking at Fig. 70 and neglecting input channels 3 and 4, the system first uses frequency modulation to modulate the subcarrier. Therefore, the first letter-group in the designation is F-M. The next modulation process in this system is the final modulation which is applied to the R-F carrier. It is also frequency modulation; therefore, the last letter group is also F-M. The complete designation for this system (still disregarding input channels 3 and 4) is F-M/F-M. The slash mark is read "ON", so the F-M/F-M when said out loud is "F-M on F-M."

Now considering again the multiplex system in Fig. 70, this time disregarding input channels 1 and 2, but observing input channels 3 and 4, the first modulation is that produced by sampling the input channels. The sampling process produces pulse amplitude modulation (PAM) (of which more will be said in the text on time division). The first letter group in the designation of the multiplex system will now be PAM. The second modulation process is that of the PAM signal frequency modulating the subcarrier. The second letter-group is therefore F-M. The next and final modulation is again F-M, therefore the complete designation for the system with respect to input channels 3 and 4 is PAM/F-M/F-M. In U. S. telemetry, phase modulation sometimes replaces frequency modulation for the final carrier. Had this been done in the case of the system in Fig. 70, F-M/P-M would apply with respect to input channels 1 and 2, and PAM/F-M/P-M would describe the system with respect to input channels 3 and 4.

Referring to the hypothetical communication system, Fig. 71A gives the spectrum resulting from FSK/AM for channels 1, 2, and 3, and SSB/AM for the speech channel. Similarly, Fig. 71B gives the spectrum for FSK/SSB and SSB/SSB. In Fig. 71C, it is impossible to distinguish the separate channels in the composite spectrum but its content is known to be FSK/FM and SSB/FM.

IV. BANDWIDTH REQUIREMENTS

Recording techniques are frequently employed in conjunction with multiplexed data. Successful recording of any information requires that the recording equipment be able to accommodate the signal bandwidth. Therefore, it is appropriate that a discussion of bandwidth be given at this time. Examples from telemetry will be used here since telemetry data is almost invariably recorded.

DOCID: 3726612

In recoverable vehicles in which data to be telemetered has a bandwidth that exceeds that allowed by the standards for multiplexed telemetry data, single or multitrack recordings are made on a miniature tape recorder contained within the vehicle itself. In non-recoverable vehicles, such as satellites, recordings are also made and, on command from the ground station, they are played back as a single channel. Another instance when recording is employed is where it is necessary for one reason or another to have a delayed transmission. This latter example may require the recording of the entire baseband which will then be transmitted at a later time on command from the ground station. Whether recording is done in the vehicle or not, transmitted telemetry signals are almost invariably recorded at the ground stations.

It is unnecessary and impractical to record the R-F signal. For this reason, recording is done after detection by the receiver. The signal remaining after detection is the baseband of subordinate carriers and their sidebands. The whole baseband may be recorded simultaneously on a wideband (video) recorder. The bandwidth required for such a recorder must be equal to or greater than the highest significant frequency in the highest subcarrier band employed. Since the highest subcarrier band employed may vary widely from one system to another, the highest possible subcarrier band (hence maximum baseband bandwidth) must be assumed in the selection of recording equipment for multiplexed signals. It is sometimes desirable to record individual subcarrier bands instead of placing the entire baseband on a multitrack tape. To do this, a channel selector is employed to separately detect each subcarrier band. The information from each band is then recorded on a narrow band recorder.

It will be remembered that when determining the bandwidth for the different forms of modulation, the highest modulating frequency always entered into the formula. Listed, these formulas are:

$BW = 2f_m$	(Amplitude Modulation)
$BW = f_m$	(Single Sideband)
$BW = 2f_m\ (M_f + 1)$, or $2(f_d + f_m)$	(Frequency Modulation)
$BW = 2f_m\ (M_p + 1)$	(Phase Modulation)

In all of these equations, f_m represents the highest modulating frequency. For this reason, after modulation the bandwidth around any subcarrier or carrier, regardless of the kind of modulation employed, is dependent upon the highest modulating frequency.

A. Baseband Bandwidth

Quite naturally, the upper frequency limit of a baseband will be

DOCID: 3726612

determined with respect to the subcarrier having the highest frequency. The actual frequency bounding the high extremity of the baseband will be the highest sideband component associated with the highest subcarrier band. For example, for band 18 the subcarrier center frequency is 70 kilocycles and the highest modulating frequency for this band is 1.05 kilocycles. The maximum deviation for this band is 75.25kc - 70kc = 5.25 kc. Substituting the necessary values into the appropriate bandwidth formula for the modulation method employed, there obtains, $BW = 2(f_d + f_m) = 2(5.25kc + 1.05kc) = 12.6kc$. This is the total bandwidth for band 18, the center frequency of which is 70 kilocycles. The highest frequency present is then equal to the center frequency plus one-half of the calculated bandwidth. This frequency, when solved by an equation, is,

$$\text{Highest Frequency} = 70 \text{ kc} + \frac{12.6 \text{ kc}}{2} = 76.3 \text{ kc}$$

In a very similar manner, the lower frequency limit of a baseband can be determined with respect to the subcarrier having the lowest frequency. The actual frequency bounding the low extremity of the baseband will be the lowest sideband component associated with the lowest subcarrier band. For example, for band 1 the subcarrier center frequency is .4 kilocycles and the highest modulating frequency to be used is six cycles. The maximum deviation is 430 cycles - 400 cycles = 30 cycles. By substitution into the bandwidth equation,

$$BW = 2(f_d + f_m) = 2\ (30 + 6) = 72 \text{ cycles/second.}$$

This is the total bandwidth for band 1, the center frequency of which is 400 cycles per second. The lower frequency limit of the baseband is equal to the center frequency minus one-half of the calculated bandwidth. This turns out to be 364 cycles per second.

The bandwidth of the baseband is the frequency range between its lower and upper frequency limits. To be practical, however, it is a good practice to think of the lower limit as extending all the way down to dc when specifying equipment for recording purposes. The bandwidth of the baseband may then be considered as extending all the way from zero frequency to the frequency of the highest subcarrier sideband component that is anticipated.

B. R-F Bandwidth

In discussing the bandwidth requirements for multiplexed signals, no new facts have been presented. The methods of computation are exactly the same as when the various forms of modulation were taught in earlier lessons. The only thing new is the way in which the methods are applied, namely, that bandwidth calculations must be made for as many modulation processes as are

DOCID: 3726612

used, and, of course, there may be variety in the kinds of modulation employed in a single system. Since modulation of the baseband subcarriers is not the final process involved, there must be additional computation if the R-F bandwidth is to be found.

It should be clear that the upper frequency limit of the baseband is the highest modulating frequency for the R-F carrier. The same formulas that applied for less complex modulating waves also apply for final modulation of an R-F carrier. Assuming an R-F carrier whose frequency is 260 megacycles per second, and the maximum prescribed frequency deviation of ± 125 kilocycles per second, the R-F bandwidth resulting from frequency modulation by the baseband previously described is,

$$BW = 2(f_d + f_m) = 2(125 + 76.3) = 402.6 \text{ kc/s.}$$

Although FM/FM telemetry has been used as an example here, bandwidth requirements for all frequency multiplex services may be solved for in a similar manner by using the proper formulas for the modulation method employed.

CONCLUSION

By allocating different carrier frequencies for each communication channel, and by controlling the spectrum bandwidth of each station's channel, radio stations are kept from interfering with one another even though they simultaneously radiate into a common medium. In exactly the same manner, multiple channels of information can be carried simultaneously in a common medium, in a single system, and without one information channel interfering with another. To do this it is only necessary to let each information channel in the system modulate a separate low frequency carrier that has a different frequency than that of any other carrier in the system. Of course, the spectrum bandwidth for the sidebands produced around each carrier must be controlled also in order that sidebands from one channel do not overlap the sidebands of another channel.

This scheme of separating information channels by the separation of carrier assignments is not any different than that employed in separating the communication channels of radio stations. Be that as it may, when this scheme is applied in a single system it is called multiplexing. By virtue of the fact that each information channel is made to share a different band in the system's frequency spectrum, it is called a frequency sharing multiplex system, or frequency division multiplexing.

To make the outputs of the several channel modulators into a single signal without interchannel interference it is only necessary to provide adequate separation between the carrier frequencies identified with each information channel and to confine the bandwidth of each channel so that there is no overlapping of spectra.

DOCID: 3726612

V. PRACTICAL EXERCISES

Questions 1 through 5 are based on the following:

A baseband has two subcarriers whose respective frequencies are 10 kc/s and 20 kc/s. These are amplitude modulated by 25 millisecond square pulses. The third harmonic is required, but a tolerance is allowed to include the fourth.

1. What is the highest assumed frequency in the baseband?

2. What is the lowest assumed frequency in the baseband?

3. What is the recording bandwidth for the baseband?

4. If this baseband is used to amplitude modulate an R-F carrier, what will the R-F bandwidth be?

5. If this baseband is used to frequency modulate a 10 megacycle R-F carrier to obtain a frequency deviation of ± 75 kc/s, what will the R-F bandwidth be?

6. What kinds of modulation may be used in a baseband?

7. Telemetry subcarrier band 17 is the highest used in the baseband for a given test. (a) What is the center frequency?

   (b) What is the frequency deviation?

8. If the highest modulating frequency used for the subcarrier in question 7 is 790 cycles per second, what is the bandwidth for this subcarrier band?

9. What is the assumed highest frequency in the baseband described in questions 7 and 8?

10. The baseband described in questions 7 - 9 is used to frequency modulate a 216 megacycle carrier and the frequency deviation is ± 125 kc/s.

(a) What is the R-F bandwidth?

(b) What is the modulation index for this final modulation?

11. (a) If the frequency deviation in problem 10 were reduced to ± 75 kc/s, but the modulation remained the same, what would the R-F bandwidth be?

(b) What is the new modulation index?

(c) Is this narrow band or wide band FM?

12. What is the purpose of a baseband?

13. If a baseband extended from DC to 2 megacycles/second, what recording bandwidth would be required?

Questions 14 through 20 are based on the following: Subcarrier band E is used in the baseband of a certain telemetry channel. This subcarrier band is carrying sampled information from a subcommutated information channel. The samples have an 833 microsecond duration, and the frame marker (sync pulse) has 2500 microseconds duration.

14. What is the assumed highest modulating frequency for the subcarrier (4th harmonic)?

15. Can this frequency modulate the subcarrier 100%, and if not, what is the highest frequency that can?

DOCID: 3726612

16. Based on the highest frequency that can fully modulate the subcarrier, what is the recording bandwidth of the baseband?

17. What is the highest frequency in the baseband?

18. What is the R-F bandwidth for the telemetry channel if the frequency deviation due to final modulation is ± 75 kc/s?

19. If the commutator referred to produces 30 samples/frame and the commutation rate is 600 samples/second, work out the frame rate then prove your answer from the tables.

20. If the information being fed to the subcommutator must be sampled at least 6 times for every cycle, what is the highest frequency that can be sampled by this system?

DOCID: 3726612

APPENDIX

In this text several tables have been reprinted from a document prepared by the Inter-Range Telemetry Working Group, and the Inter-Range Instrumentation Group, White Sands Proving Ground, New Mexico. The work is identified as IRIG Document No. 103-56, Telemetry Standards for Guided Missiles. Part I, deals with the FM/FM and FM/PM standards, which apply to frequency division multiplexing used with telemetry in the United States. Part I with its published revisions and addenda have been reproduced below for the benefit of those seeking source information.

PART I. FM/FM OR FM/PM STANDARD

1. GENERAL

These telemetry systems are of the frequency division multiplex type. That is, an r-f carrier is modulated by a group of subcarriers, each of a different frequency. The subcarriers are frequency modulated in a manner determined by the intelligence it is desired to transmit. One or more of the subcarriers may be modulated by a time division multiplex scheme (commutation) in order to increase considerably the number of individual data channels available in the system. The modulation of the r-f carrier may be by either of two methods: frequency modulation or phase modulation.

2. SUBCARRIER BANDS

Eighteen standard subcarrier band center frequencies with accompanying information on frequency deviation and nominal intelligence frequency response are specified in Table III. It is intended that the standard FM/FM receiving stations at the test ranges be capable of simultaneously demodulating a minimum of any twelve of these subcarrier signals.

All test ranges may not presently have demodulating equipment for all the subcarrier bands specified in Table III. Therefore, it is important that the missile or aircraft contractor expecting to require telemetry receiving services from a test range coordinate his particular subcarrier band usage with the test range at the time he first starts to plan the telemetry transmitting system. A similar recommendation applies if the contractor desires more than twelve subcarrier signals (multiplexed on a single r-f carrier) to be simultaneously demodulated.

The upper five bands, when used with a ± 15 percent frequency deviation, are listed as being "optional", and it is intended that each range provide the equipment for demodulating these optional bands as the need arises.

DOCID: 3726612

The normal frequency response listed for each band is computed on a basis of maximum deviation and a deviation ratio of five, and it is intended that the standard receiving station be capable of demodulating data with these frequency responses. However, it should be remembered that the actual frequency response obtainable is dependent on many things, such as the actual deviation used, the characteristics of filters, etc. The primary reason for specifying a frequency response is to insure that elements in the receiving station such as filters and recording oscillographs do not limit the frequency responses shown in Table III.

Deviation ratios as low as one or less may be used, but low signal-to-noise ratios and possible increased harmonic distortion and cross-talk must be expected.

DOCID: 3726612

TABLE III. SUBCARRIER BANDS

Band	Center Frequency (cps)	Lower Limit (cps)	Upper Limit (cps)	Maximum Deviation (percent)	Frequency Response* (cps)
1	400	370	430	7.5	6.0
2	560	518	602	7.5	8.4
3	730	675	785	7.5	11
4	960	888	1,032	7.5	14
5	1,300	1,202	1,398	7.5	20
6	1,700	1,572	1,828	7.5	25
7	2,300	2,127	2,473	7.5	35
8	3,000	2,775	3,225	7.5	45
9	3,900	3,607	4,193	7.5	59
10	5,400	4,995	5,805	7.5	81
11	7,350	6,799	7,901	7.5	110
12	10,500	9,712	11,288	7.5	160
13	14,500	13,412	15,588	7.5	220
14	22,000	20,350	23,650	7.5	330
15	30,000	27,750	32,250	7.5	450
16	40,000	37,000	43,000	7.5	600
17	52,500	48,560	56,440	7.5	790
18	70,000	64,750	75,250	7.5	1,050
A**	22,000	18,700	25,300	15	660
B	30,000	25,500	34.500	15	900
C	40,000	34,000	46,000	15	1,200
D	52,500	44,620	60,380	15	1,600
E	70,000	59,500	80,500	15	2,100

*Frequency response given is based on maximum deviation and a deviation ratio of five (see text for discussion).

**Bands A through E are optional and may be used by omitting adjacent bands as follows

Band Used	Omit Bands
A	15 and B
B	14, 16, A and C
C	15, 17, B and D
D	16, 18, C and E
E	17 and D

NOTE: In the process of magnetic tape recording of the above listed subcarrriers at a receiving station, provision may also be made to record a tape speed control tone and tape speed error compensation signals as specified in the Magnetic Tape Recorder/Reproducer Standards (to be issued soon). The speed control tone frequency is 17,000 cps. The compensation signal frequencies are 50,000 cps and 100,000 cps at tape speeds of 30 and 60 inches per second respectively.

DOCID: 3726612

The eighteen bands were chosen to make the best use of present equipment and the frequency spectrum. There is a ratio of approximately 1.3:1 between center frequencies of adjacent bands except between 14.5 kilocycles and 22 kilocycles, where a larger gap was left to provide for a possible compensation tone if magnetic tape recording is used. The deviation has been kept at ± 7.5 percent for all bands with the option of ± 15 percent deviation on the five higher bands to provide for transmission of higher frequency data. When this option is exercised on any of these five bands certain adjacent bands cannot be used, as listed in the footnote to table III.

It is likely that certain applications will make amplitude pre-emphasis of some subcarrier signals desirable, and it is recommended that the ground equipment be capable of accommodating this pre-emphasized signal. A de-emphasis capability of up to 9/db per octave may be required.

3. COMMUTATION

Commutation (time division multiplexing) may be used in one or more subcarrier bands. A nearly limitless variety of commutation schemes could be devised, but a few relatively simple methods will satisfy most telemetry needs. The specifications listed below for commutation were chosen to give maximum flexibility consistent with presently available equipment and techniques, and it is intended that, in order to limit the varieties which must be handled at test ranges, the following restrictions on commutation be observed:

Certain commutation rates shall not be exceeded on each subcarrier. In all cases where automatic channel separation is desired, the commutated signal shall have a definite pattern and rates as described below.

It is required that the standard receiving station be capable of accepting commutated data with the following specifications:

a. Unseparated Data: When automatic channel separation of a commutated subcarrier (decommutation) is not used, the commutated data may be recorded directly, i.e., from each subcarrier demodulator the data shall appear in time sequence. Commutation rates should not exceed the values listed in Table IV. These rates provide samples sufficiently long to permit optical interpretation and to maintain cross-talk at tolerable levels.

(1) Conservative Values: The sample durations listed in the conservative value column of Table IV are chosen sufficiently long that the reading error during the last half of the sample, due to the switching transient, will not exceed approximately one-half percent of the quiescent value. These values apply when the recording equipment is limited by a low-pass filter whose cut off frequency is equal to the specified frequency response of that subcarrier band as given in Table III.

DOCID: 3726612

TABLE IV. COMMUTATION RATES - UNSEPARATED DATA

Band	Center Frequency (cps)	Sample Duration (Milliseconds) Conservative Values	Sample Duration (Milliseconds) Minimum Values	Commutation Rate* (samples per second) Conservative Values	Commutation Rate* (samples per second) Maximum Values
1	400	670	170	1.5	6.0
2	560	480	120	2.1	8.4
3	730	370	91	2.7	11
4	960	280	70	3.6	14
5	1,300	210	51	4.9	20
6	1,700	160	39	6.4	25
7	2,300	120	29	8.6	35
8	3,000	89	22	11	45
9	3,900	68	17	15	59
10	5,400	49	12	20	81
11	7,350	36	9.1	28	110
12	10,500	25	6.4	39	160
13	14,500	18	4.6	55	220
14	22,000	12	3.0	83	330
15	30,000	8.9	2.2	110	450
16	40,000	6.7	1.7	150	600
17	52,500	5.1	1.3	200	790
18	70,000	3.8	0.95	260	1,050
A	22,000	6.1	1.5	170	660
B	30,000	4.4	1.1	230	900
C	40,000	3.3	0.83	300	1,200
D	52,500	2.5	0.63	390	1,600
E	70,000	1.9	0.48	530	2,100

*Frame rate times the number of samples per frame. This assures no lost time between samples. Multiply this value by the duty cycle for the actual values.

(2) Minimum Values: The sample lengths listed in the minimum value column of Table IV provide a sample that is one-fourth the length listed in the conservative value column. These values are to be used only when the usual low-pass filters for the deviation ratio of five have been replaced with units having cutoff frequencies about four times the recommended frequency response. The removal of the regular low-pass filter will permit the recording of a considerable

DOCID: 3726612

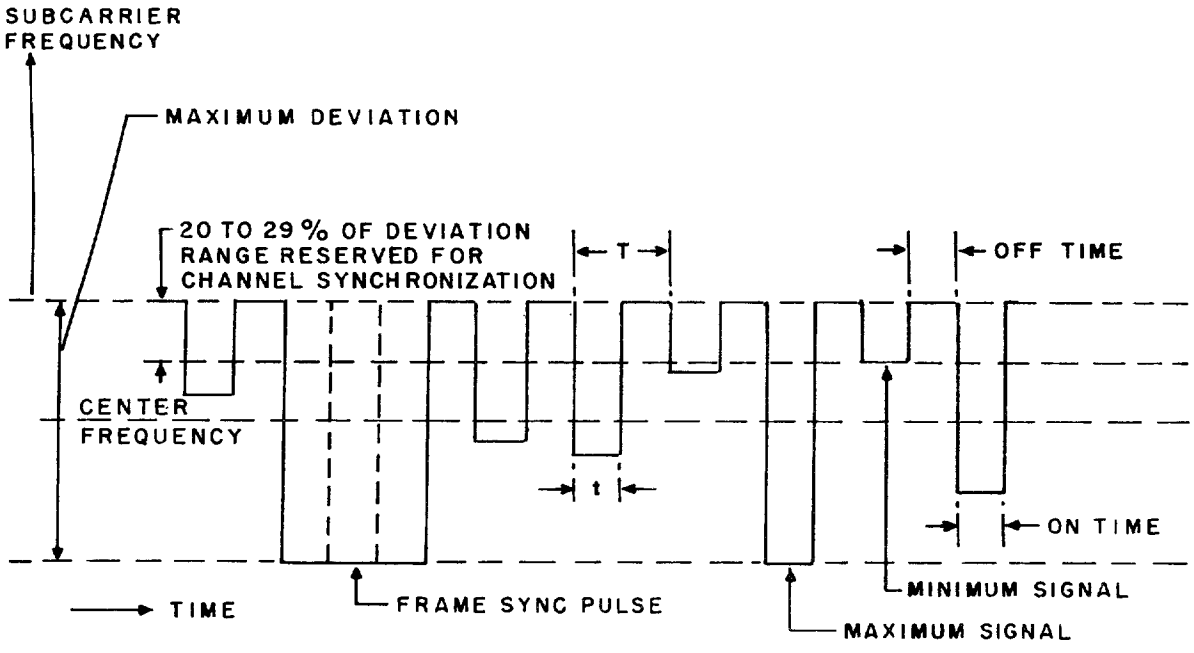
amount of spurious signals. The reading accuracy of samples recorded in this manner will be considerably reduced.

b. Separated Data: Where required, automatic channel separation (decommutation) equipment shall be provided in the receiving station to process commutated signals that conform to the following characteristics:

(1) The total number of samples per frame (number of segments of a mechanical commutator) and the frame rates shall be one of the combinations shown in Table V. If a higher repetition rate is required for certain information, two or more samples per frame (equally spaced in time) can be used to represent one telemetered function at the expense of the total number of information channels.

(2) The commutation pattern, in the subcarrier frequency vs time domain, shall be as shown in Figure 72.

(3) A frame synchronizing pulse of duration equal to two "on" periods plus one "off" period shall be provided once every frame, as shown in Figure 72.



$\frac{t}{T}$ = DUTY CYCLE

FIGURE 72.

DOCID: 3726612

(4) The commutator speed (or frame rate) shall not vary more than + 5 to -15 percent from the nominal values given in Table V.

(5) When airborne gating is used, the duration of each information pulse shall be within the bounds of 47 to 53 percent of nominal channel period and shall be stable to plus or minus one-half percent of nominal channel period. It should be noted that stability of information pulse duration often affects system accuracy. Changes in the channel period due to changes in frame rate must be absorbed in the "off" time.

(6) When airborne gating is not used, it may be necessary to use ground gating and the ratio of information pulse duration to nominal channel period shall be within the bounds of 60 to 75 percent.

TABLE V. COMMUTATION SPECIFICATIONS FOR AUTOMATIC DECOMMUTATION

18	5	90	14,500
18	10	180	22,000 (±15%) or 30,000 (±7.5%)
18	25	450	30,000 (±15%) or 70,000 (±7.5%)
30	2.5	75	10,500
30	5	150	22,000 (±7.5%)
30	10	300	22,000 (±15%) or 40,000 (±7.5%)
30	20	600	40,000 (±15%)
30	30	900	70,000 (±15%)

*The number of samples per frame available to carry information is two less than the number indicated, because the equivalent of two samples is used in generating the frame synchronizing pulse.

**Frame rate times number of samples per frame.

4. AUTOMATIC CORRECTION OF SUBCARRIER DRIFT AND GAIN VARIATION

Signals multiplexed with data subcarriers may be transmitted for the purpose of providing automatic correction of subcarrier and/or discriminator

drift and gain. Automatic Drift and Gain Correction signals shall conform to the following standards:

Calibration Command

A standard IRIG subcarrier band shall be employed for Calibration Commands. Exceptions shall be coordinated with the Test Range concerned.

Calibration Interval

Calibrations shall be performed as often as possible consistent with data requirements. Calibration intervals shall be programmed so as to avoid interference with critical portions of the test.

Calibration Sequence

The Command Subcarrier Band shall normally operate at its center frequency $\pm$ 0.75%. The calibration sequence shall consist of Center Frequency Correction followed by Gain Correction signals as follows:

a. Center Frequency Correction

(1) Command Subcarrier Band

The command to correct for center frequency shall consist of a displacement of the Command Subcarrier upward in frequency to 6.75% f_c. The command shall last for 50% of the total calibration interval.

(2) Data Subcarrier Bands

Data bands shall be connected to signals representing nominal center frequency 200 milliseconds (+ 50, -0 milliseconds) after the command band shifts to its upper frequency. Data bands shall remain at center frequency 200 milliseconds after the command band shifts to its lower frequency.

b. Gain Correction

(1) Command Subcarrier Band

The command to correct for gain changes shall consist of a displacement of the command subcarrier from its upper

frequency to a frequency of -6.75% f_c ± 0.75% f_c. This command shall last for 50% of the total calibration interval after which time the command subcarrier shall return to center frequency.

c. Calibration Sequence Duration

For a correction capability of ± 10% of bandwidth per calibration cycle, the Center Frequency Correction and Gain Correction signals shall each be supplied for 2-1/2 seconds. Where maximum corrections per calibrate cycle are not required, calibrate intervals may be correspondingly reduced.

DOCID: 3726612

CHAPTER 7

TIME DIVISION MULTIPLEXING

Telemetry is one of the major applications of time division multiplexing. For this reason it has been thought convenient to center the discussion on multiplexing by time sharing around telemetry systems. It should be made clear, however, that the association is not intended to infer that telemetry is the only significant application for this type of multiplexing. Indeed, an effort has been made to keep the discussion broad enough that the student can extend the examples cited for telemetry to include other types of time multiplex systems as well.

The analysis of the type of signal produced by time division multiplexing should give the student an appreciation for the pulse analysis techniques taught earlier in this series.

I. PULSE MODULATION

The basic requirement of telemetry and the root meaning of the term itself is "to measure at a distance." Therefore, the varieties of information processed in telemetry, although quite varied in nature, are in almost every instance types by which some measurement can be determined. The information at its source is usually a physical or mechanical variation. In order to transmit these variations electrically, they must be transformed into electrical waves whose characteristic variations are proportional to those of the information wave produced by the source. The overall devices used to produce the electrical waves from the physical or mechanical variations are called end-instruments or tranducers. The output waves from the end-instruments are generally quasistatic and usually continuous rather than discrete.

For time division multiplexing, the outputs of the end-instrument must be caused to modulate carriers, not consisting of sinewaves but rather of pulses. This may come as somewhat of a surprise to the student whose previous acquaintance with modulation has been limited to that of sinewave carriers. The pulse carriers used in time division multiplexing have the same significance as the subordinate carriers used in frequency division multiplexing. Just as frequency division multiplexes by stacking sinusoidal subcarriers of different frequencies, so time division multiplexes by interleaving pulse subcarriers by making use of the interval between pulses. It is appropriate, therefore, to consider some of the basics of the modulation of pulse carriers before going into a discussion of time division multiplexing itself.

Basically there are two classifications of pulse modulation: the first and

DOCID: 3726612

simplest is pulse amplitude modulation and the second classification, which takes several forms, is pulse time modulation. Concerning the former, pulse amplitude modulation (PAM) is simply the amplitude modulation of a pulse carrier. Concerning the latter, pulse time modulation (PTM) is "modulation in which the time of occurrence of some characteristic of a pulse carrier is varied from the unmodulated value" (51 IRE 20.S1; 53 IRE 11.S1). Pulse time modulation is a general term which applies equally to several types of pulse modulation, some of which are pulse duration modulation (PDM), pulse position modulation (PPM), and pulse frequency modulation (PFM). (It is noted that the term pulse time modulation once referred to pulse position modulation exclusively.)

Some question might arise as to the place held by pulse code modulation (PCM). This system is not a modulation at all, but simply a code consisting of pulse combinations. Hence, it has no specific place in either pulse amplitude modulation or pulse time modulation. A pulse code may consist of variations in order, number, spacing, amplitude, or other characteristics. The term, pulse code modulation is, therefore, a very general one and requires some additional specification to indicate just what form the pulse code takes. Two additional designations that belong to the category of pulse code modulation are pulse number modulation (PNM) and pulse count modulation (abbreviated PCM, also).

A. Pulse Amplitude Modulation

Pulse amplitude modulation is the pulse counterpart of the amplitude modulation which utilizes sinusoidal carriers. Two methods of obtaining an amplitude modulated pulse train are described.

1. PAM from a Pulse Carrier

This method of producing pulse amplitude modulation involves generating a train of rectangular pulses whose pulse duration and pulse repetition frequency are fixed. This pulse train will bear the same relationship to a wave of modulating intelligence as a sinewave carrier does in standard amplitude modulation. This pulse train can, therefore, be called a pulse carrier in contrast to a sinewave carrier. A typical pulse carrier is illustrated in Figure 73A.

Prior to modulation the amplitudes of all the pulses in the carrier are at the same level. In a special modulator designed for this type of waveform, the information wave (Fig. 73B) is caused to produce a pulse wave train whose pulses have variations in amplitude corresponding to the amplitude changes in the information wave. Observe Figure 73C.

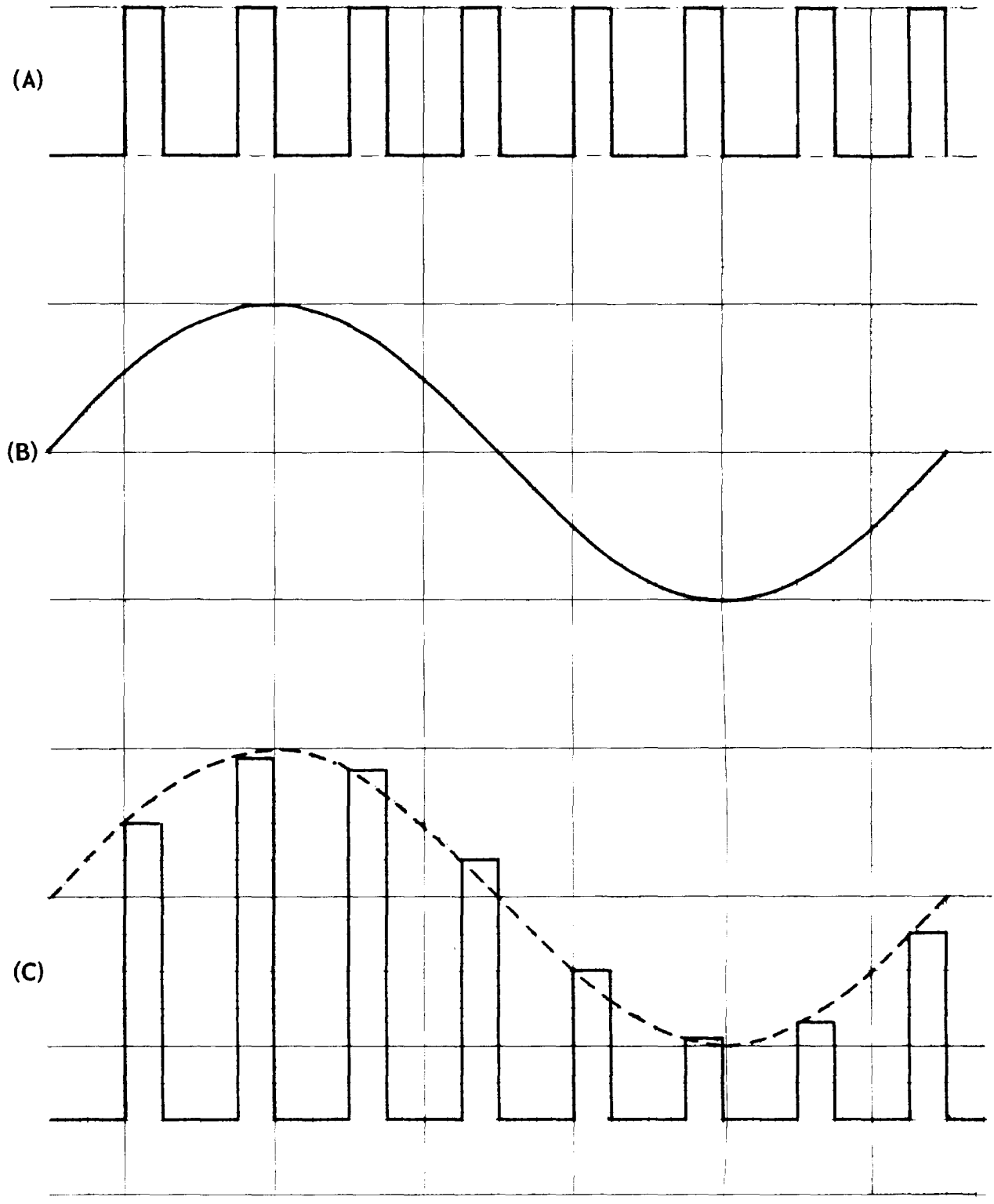


FIGURE 73. AMPLITUDE MODULATION OF A RECTANGULAR WAVE TRAIN.

DOCID: 3726612

2. PAM from Sampling

One of the simpler means of producing PAM is that of sampling. A sample according to Webster is, "a part of anything . . . shown as evidence of the quality of the whole." This definition is not quite broad enough to be applied directly to sampling theory in communication and telemetry, but does give a hint of the same basic idea. The big difference lies in the fact that in electronic sampling the whole can be recovered from a number of samples without substantial loss of quality, if the original wave has been sampled properly.

It is a thought that may seem strange at first, but information waves, whether sinusoidal or complex, can be sampled. By sampling, it is meant that the amplitude of the information wave, is measured at discrete instants of time and at fixed regular intervals along the wave. Each such measurement constitutes a sample. The relationship of a wave and the samples taken of it is shown in Figures 74A and 74B. Provided the time interval between the amplitude samples is sufficiently short, all the information in the original wave can be recovered.

It will be noticed that the pulse train obtained from sampling is almost identical to that obtained by amplitude modulating a pulse carrier. The basic difference in appearance is that the pulses obtained by modulating a pulse carrier are flat-topped whereas those obtained by direct sampling are curve-topped corresponding to the curvature of the modulating wave during the time when the sample was taken. If the pulse duration in both instances is made extremely short, there will be no difference in the shape of the pulse trains acquired by the two methods.

3. Evaluation of PAM

Whether obtained by direct sampling or by the process of modulating a pulse carrier, the resultant pulses may be properly regarded as samples (for all practical purposes) of the amplitude of the modulating wave. By transferring the information from the modulating wave to the sequence of samples, the information is presented intermittently rather than continuously. Sampling, then, takes continuous information and converts it into discrete information in the form of voltage pulses. The amplitude of each pulse represents the instantaneous amplitude of the original wave at the time when the sample was taken.

DOCID: 3726612

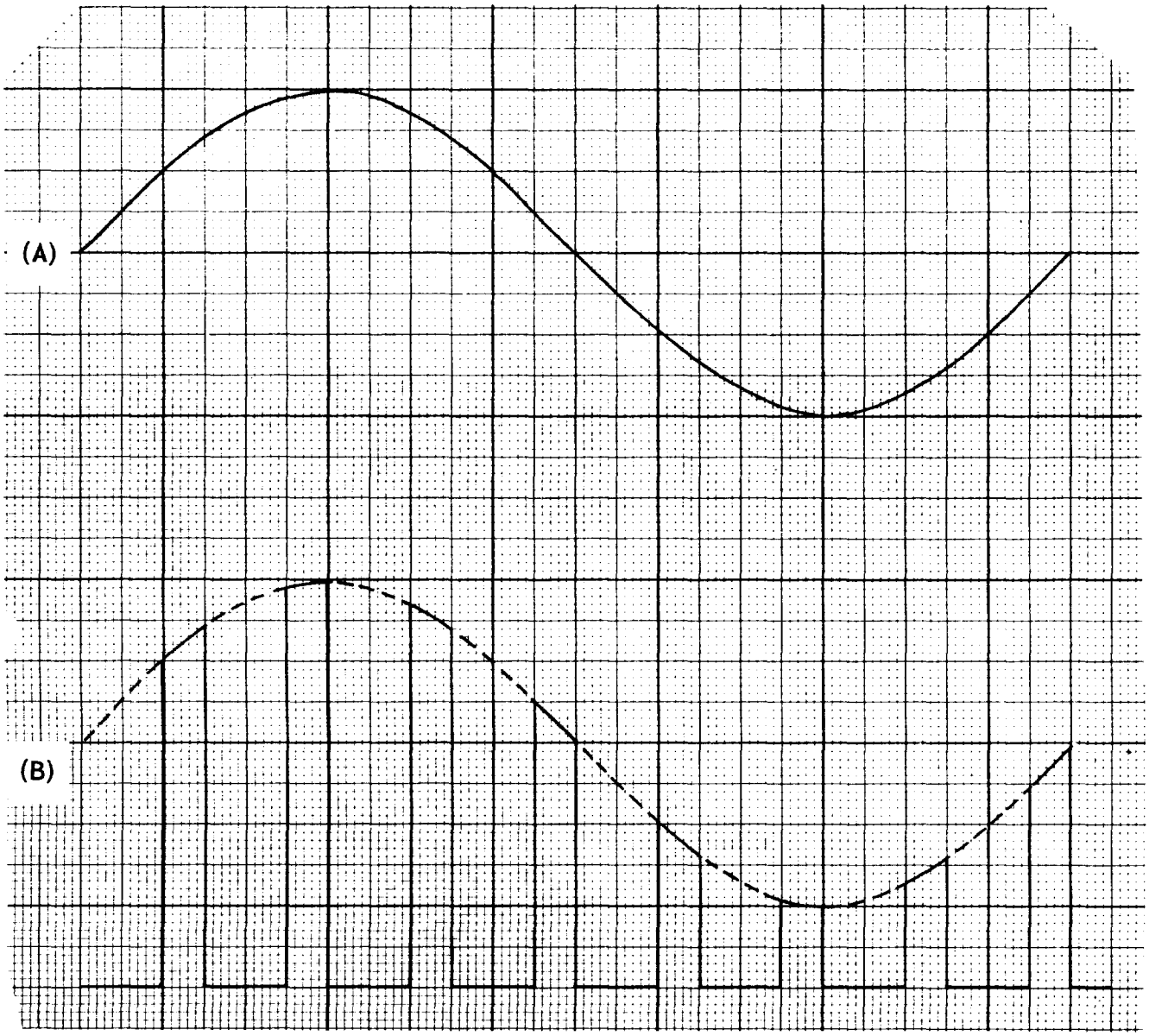


FIGURE 74. PULSE AMPLITUDE MODULATION BY SAMPLING.

DOCID: 3726612

To recover perfectly all the information and quality of the modulating wave, the pulse duration of the rectangular samples would have to be infinitely small--a thing which cannot be achieved in any actual system. It is also necessary that the PRF of the samples be at least twice the highest frequency to be sampled. A lesser PRF will allow sidebands from around adjacent PRF harmonics to overlap and thus interfere with one another. In actual practice, a minimum of slightly more than two samples per cycle of the highest modulating frequency are taken. An example from communication practice is where high quality voice transmissions are made by taking 2.4 samples for the highest voice frequency (3333.3 cps) to be transmitted. For this particular example the number of pulses per second would be,

$$\frac{2.4 \text{ samples}}{\cancel{\text{Cycle}}} \times \frac{3333.3 \cancel{\text{cycles}}}{\text{Second}} = \frac{7999.2 \text{ samples}}{\text{Second}}$$

(In actual usage this is 8000 samples/second.) To take a further example, one applicable to telemetry, what is the minimum sampling rate that can probably be used for a quasistatic variation whose bandwidth is 10 cycles per second? This is solved quite easily as follows,

$$\frac{2 \text{ samples}}{\cancel{\text{Cycle}}} \times \frac{10 \cancel{\text{cycles}}}{\text{Second}} = \frac{20 \text{ samples}}{\text{Second}}$$

The time allotted to each sample in this latter example would be:

$$\frac{\cancel{\text{Second}}}{20 \text{ samples}} \times \frac{1000 \text{ milliseconds}}{\cancel{\text{Second}}} = \frac{50 \text{ milliseconds}}{\text{Sample}}$$

This time is the period of the PAM wave train. The duration of the samples could be as much as one-half the period (yielding squarewaves) but in practice it is much less in order to permit multiplexing of other pulses between the samples.

There are times when telemetry requires extreme accuracy, This can be obtained by a high sampling rate. When accuracy cannot tolerate more than .1% error, as many as 6 samples must be taken for each cycle of the highest modulating frequency.

DOCID: 3726612

4. The PAM Spectrum

The spectrum of an amplitude modulated rectangular pulse carrier is basically that of a similar unmodulated rectangular pulse train. In both instances the fundamental frequency and the required harmonics for the particular pulse train are present. These obey the (Sine X) /X relationship discussed earlier in this course. When the rectangular pulse carrier is subjected to modulation, each Fourier component (harmonic) in the pulse train becomes, as it were, a separate sinewave carrier in its own right. The effect of the modulating process, then, is to amplitude modulate these harmonics so as to produce sidebands around each as illustrated in Figure 75.

Depending upon the method used, it is possible to have bi-directional (double polarity) PAM pulses instead of the unidirectional (single polarity) pulses that have occupied the discussion so far. Figure 76 illustrates the bidirectional pulses and their resultant spectrum. The basic difference in the spectrum is that the PRF and its harmonics disappear but the sidebands remain and are centered around the points in the spectrum where the harmonics would have been.

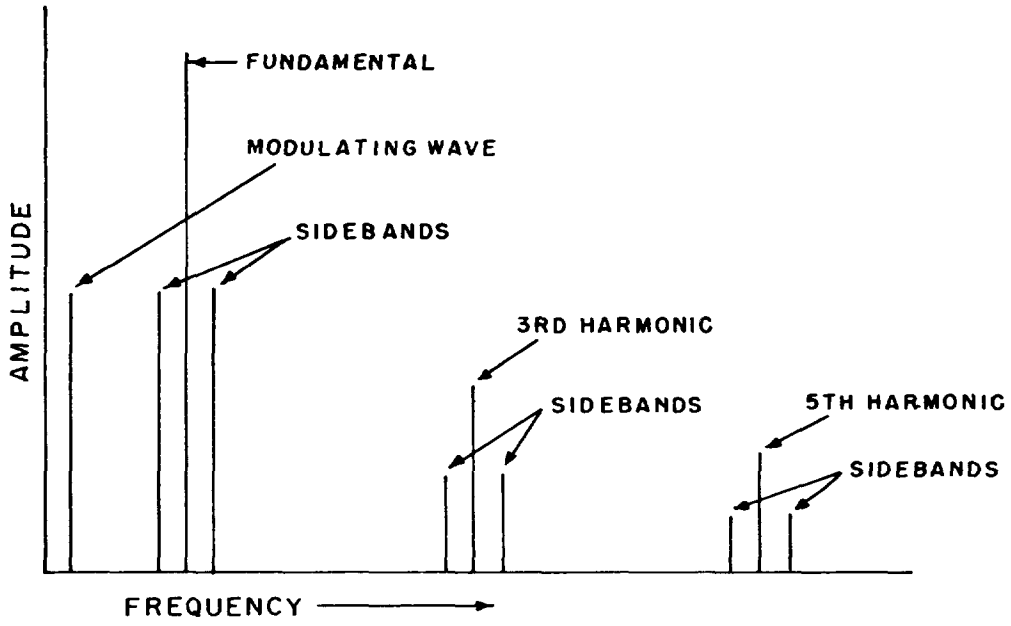


FIGURE 75. SPECTRUM FOR AMPLITUDE MODULATED SQUAREWAVE.

DOCID: 3726612

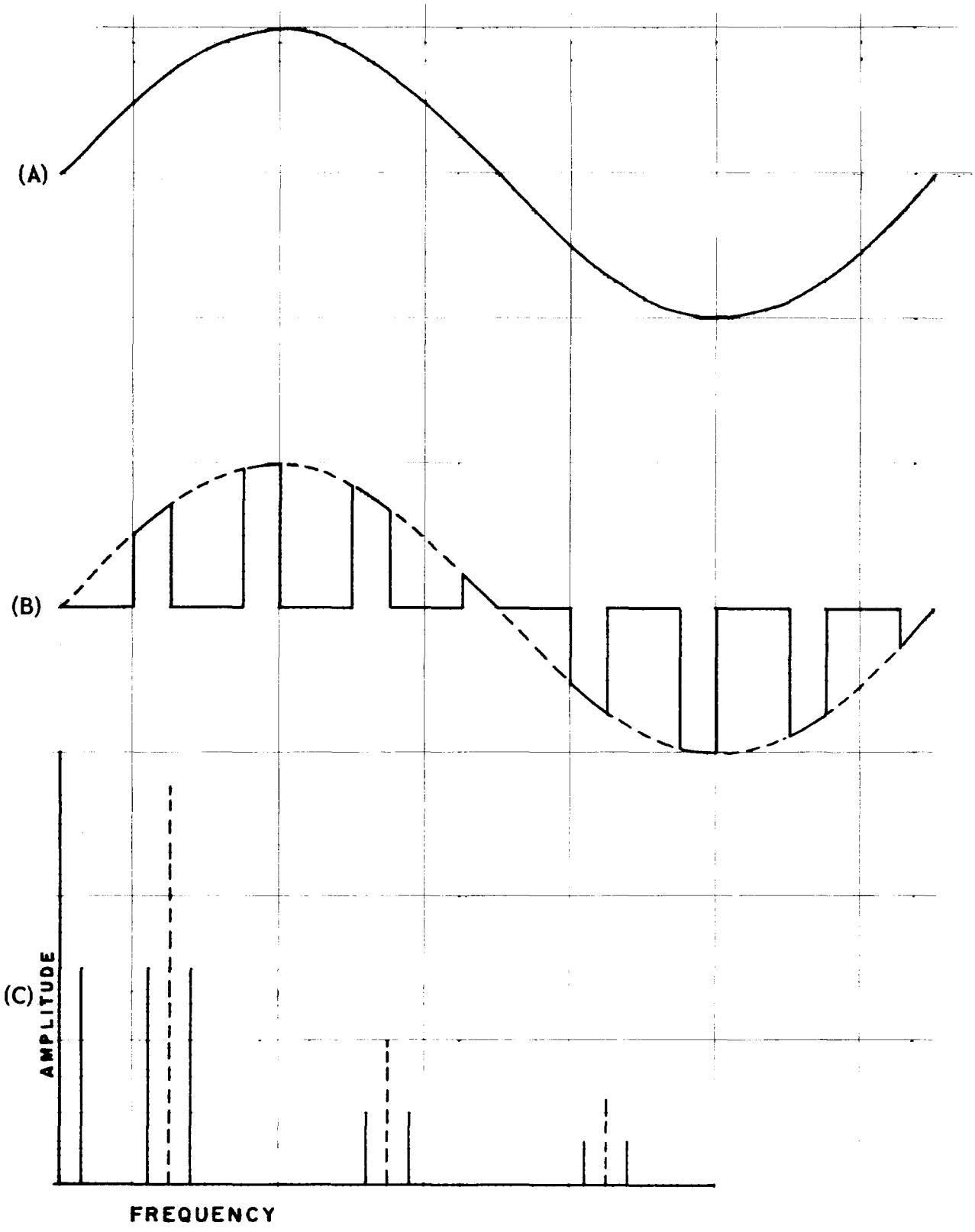


FIGURE 76. DOUBLE POLARITY SAMPLING..

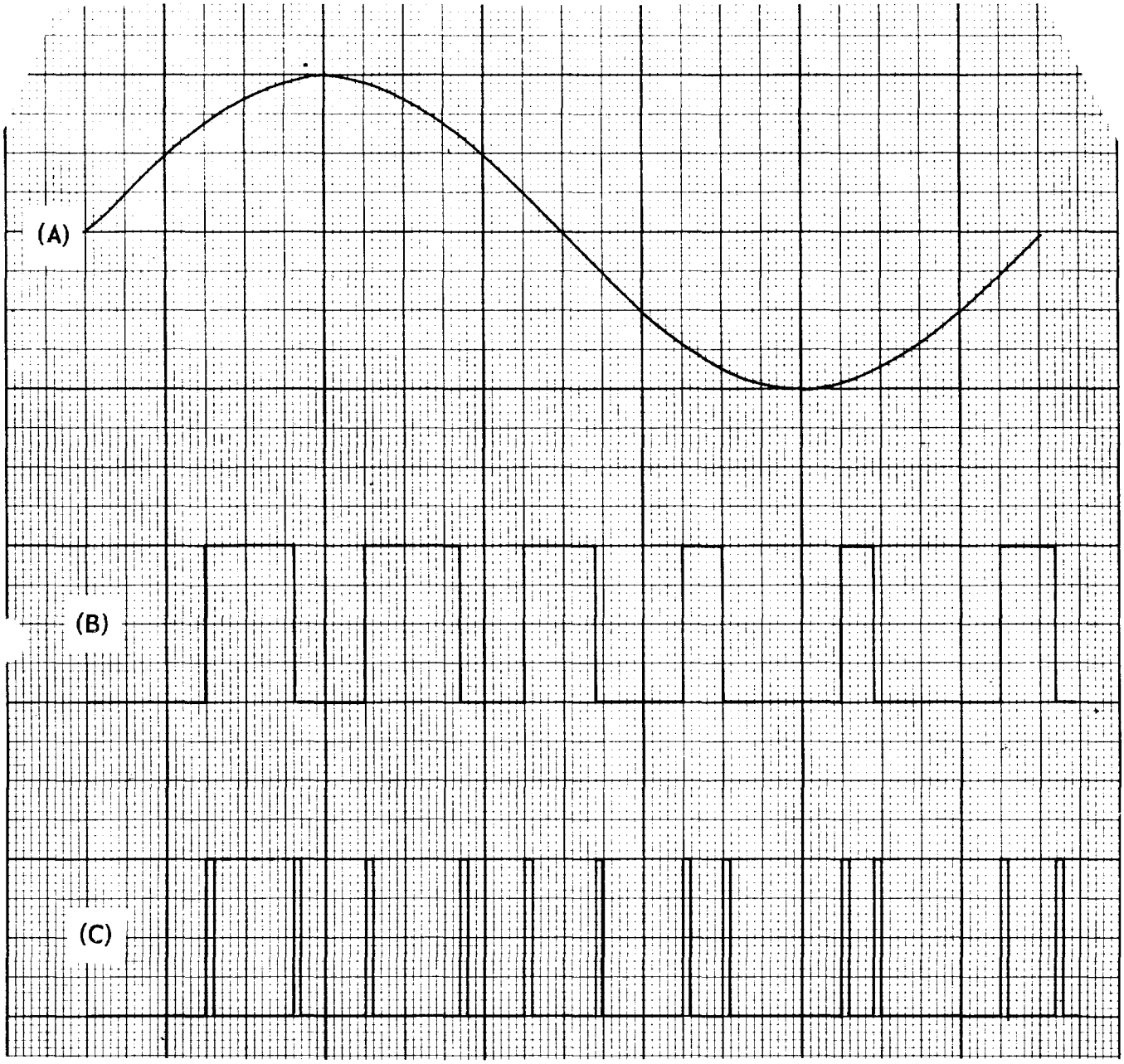


FIGURE 77. PDM AND PPM.

DOCID: 3726612

B. Pulse Duration Modulation

This is one of the forms of pulse time modulation, i.e., a modulation in which the time of occurrence of one of the pulse characteristics is caused to vary in accordance with variations in the modulating wave. In PDM the time duration of the modulated pulses is varied in proportion to the amplitude of the modulating signal. This may be accomplished by varying the leading edge, trailing edge, or both edges of the modulated pulse. Figures 77A and B illustrate pulse duration modulation where the leading edge is fixed but the trailing edge varies in proportion to the instantaneous amplitude of the modulating wave. PDM is sometimes called pulse length modulation (PLM) or pulse width modulation (PWM).

1. Methods of Obtaining PDM

There are two basic methods whereby PDM pulses may be obtained. The one method causes the modulating wave (the wave being sampled) to alter the level of an accurately timed sawtooth wave. This is done by linearly mixing the two waves. The sawtooth wave which now rides at a varying level determined by the amplitude of the modulating wave, is then clipped at some fixed amplitude so as to produce a nearly rectangular wave the duration of which is proportional to the modulating amplitude. This method of producing PDM is called natural sampling. The second method for obtaining PDM pulses is better adaptable to the time multiplex methods used in present day telemetry. In this method the PAM wave representing samples of the modulating wave are fed to a keyer or converter. By means of the keyer, the PAM samples are used to generate new pulses of variable duration. This method is called uniform sampling because the samples are space uniformly in time. Uniform sampling converts PAM to PDM.

2. Evaluation of PDM

It was found that in PAM the information from the modulating wave was transferred to the variations in pulse amplitude. Similarly, in PDM the information from the modulating wave is transferred to the variations in pulse duration. Since the information is contained only in the variation in the edge of the pulse, the balance of the pulse which contains no information represents a useless expenditure of energy. This waste of energy is the main disadvantage of PDM. The main advantage is the great reduction in noise and interference. This outweighs the disadvantage.

DOCID: 3726612

As with PAM, considerable space must be left between the PDM pulses if time multiplexing of channels is to be possible. For PDM, however, an even greater interval is necessary in order for the duration of the pulses to vary to their maximum limit without adjacent pulses overlapping.

3. The PDM Spectrum

No attempt will be made here to evaluate the individual spectral components of a PDM wave train. It will be better for purposes of the present discussion to treat the spectrum on the basis of a single transient rectangular pulse whose duration is equal to that of the shortest element in the PDM wave train.

C. Pulse Position Modulation

A second form of pulse time modulation is pulse position modulation (PPM), sometimes called pulse phase modulation. In the past the term pulse time modulation was also used exclusively to designate pulse position modulation but that usage is no longer acceptable.

As the name implies, with pulse position modulation the time position of the pulses making up the modulated pulse train is varied in proportion to the amplitude of the modulating signal. Figure 77C illustrates pulse position modulation.

1. Methods of Obtaining PPM

It was stated previously that the information in a PDM wave is contained in the varying edge of the modulated pulses, and that the balance of the PDM pulses was wasted energy since it contains no information. The question might arise as to what would happen if the useless portion of the PDM pulses were removed and just the varying edges were retained. To follow this suggestion is to produce PPM, a form of pulse modulation that is identical to PDM except that the variable edge containing the information is replaced by a variable pulse of short duration. As compared to PDM, the equivalent PPM is a great saving in power that would otherwise be wasted.

Several methods are available for obtaining PPM pulses, but perhaps the best for the present discussion is to consider them as being derived directly from PDM pulses which in turn may be generated by either natural or uniform sampling. This type of PPM generation produces a fixed pulse for the fixed edge of the

DOCID: 3726612

PDM pulse and a variable pulse for the variable edge of the PDM pulse. The fixed pulse is very useful because it provides a time reference from which the change in position of the variable pulse can be measured. In a multiplex system the fixed reference pulses would serve the added purpose of channel markers. In short, the fixed pulse is a time reference, and the variable pulse is an information pulse.

2. Evaluation of PPM

The decided advantage of PPM over PDM has already been pointed out. It is that there is a great saving of power by eliminating that portion of a PDM pulse which can serve no useful purpose. Like PDM, PPM has the advantage over PAM in that there is the great reduction in noise. In this sense the various forms of pulse time modulation have an advantage over pulse amplitude modulation that is similar to the advantage that angle modulation has over amplitude modulation.

There is also a disadvantage that is quite significant. This disadvantage rests in the fact that PPM pulses are very short in duration. It will be remembered that short bursts of energy require a much greater bandwidth than wide pulses.

The same allowance must be made for PPM as with PDM in providing a sufficient time interval between pulses so that time multiplexed pulses from one channel will not overlap or interfere with the pulses of another channel.

3. The PPM Spectrum

The bandwidth requirement for the PPM wave train will be based upon the requirement for a single transient rectangular pulse which has like duration as the PPM pulses. No attempt will be made in this writing to evaluate individual spectral components.

D. Pulse Frequency Modulation

A third form of pulse time modulation has been enumerated in this chapter; it is pulse frequency modulation (PFM). It is a type of pulse modulation wherein the PRF of the modulated pulses varies in proportion to the amplitude of the modulating wave. It is not used in multiplexing and is not applicable to time division because due to the changing PRF, freedom from adjacent channel interference cannot be realized. For this reason no further discussion will be devoted to it.

DOCID: 3726612

E. Pulse Code Modulation

This, in itself, is not a modulation. Pulse code modulation is an application of any one of the previously mentioned pulse modulation schemes in which pulse amplitude, pulse duration, pulse position or the number of pulses in a group represent a code symbol. The code symbol, in turn, may represent a sample interpreted in terms of a quantum level. An interesting and concise discussion of PCM and quantum levels is given in Methods of Modulation. The discussion includes an actual PCM system employed in communication. The student is strongly urged to consult this reference.

II. PULSE TRAIN SYNTHESIS BY COMMUTATION (TIME DIVISION MULTIPLEXING)

The whole reason for multiplexing in any form is to simultaneously send a number of message bearing signals over a common transmission channel. In time division multiplexing, this is accomplished by the synthesis of a pulse train divided into frames, an equal period of time being allotted to each frame. Each such frame must correspond to one revolution of a mechanical (or electronic) commutator. The channels are sequential within each frame, and each channel is represented by one sample within each frame. (An exception where a channel is represented by more or less than one sample per frame is when either strapping or subcommutation is employed [covered later] .) The time required for a frame is dependent upon (1) the number of channels of information, (2) the required sampling rate for the channel carrying the highest frequency, (3) the type of pulse modulation to be used, and (4) the time required for synchronization of the system.

In order to synchronize the demultiplexing system in a telemetry receiver to the multiplexing system in the telemetry transmitter, it is necessary to provide frame markers. These are employed to indicate when a frame begins or ends. Although a frame marker does not contain any information, it does occupy channel space and for this reason it is considered as a channel in the overall multiplex system. For example, a 45 channel time multiplex system in the U.S. Telemetry Standards has only 43 channels of information, a portion of each frame equal to two channels being occupied by the frame marker. The frame marker must have some distinguishing features to prevent it being confounded with the information pulses. This distinguishing feature may be (1) multiple pulses as two, three, or four pulses distinctively spaced, (2) a pulse or pulses of greater or lesser amplitude than the rest of the train, (3) a pulse of very long or very short duration compared to the information pulses, (4) an absence of pulses, or (5) a coded symbol. In U. S. Telemetry, the absence of two pulses is used in pure time multiplex systems and where time division is used in a frequency multiplex system a pulse equal to the width of two information pulses plus the off-time between them (i.e., the time occupied by two successive channels) constitutes the frame markers.

DOCID: 3726612

A. Commutation

Commutators are of various types, both mechanical and electronic; but whatever the form, all commutators are basically a multiposition switch designed to make a continuous selection of its position in a fixed time sequence. In the present discussion a mechanical analogy will serve to illustrate the purpose and operation of commutators of all types.

A mechanical commutator is a rotary switch with a number of poles or switch positions. Selection of the positions or poles is made by means of a movable electrical contact called a brush. The brush, in turn, is mounted on a rotor which is made to rotate by means of a constant speed motor. By this scheme, the brush is caused to make contact with each of the poles in succession. The brush, therefore, makes and breaks contact with each commutator pole once for each complete revolution of the rotor. Normally, each pole of the commutator is connected to a separate end-instrument (exception to this will be noted later) each representing a separate channel of information from a different information source. Each pole of the commutator, therefore, is the input to the telemetry time multiplex system for a separate information channel.

1. Commutation - A Means of Sampling

Commutation is a means of obtaining true samples of the continuous information supplied by the end-instrument. (Of course, due to the commutation there is the automatic conversion of the continuous information into discrete information, but this is more a characteristic of sampling itself than of the means employed to accomplish sampling.) During the time that the commutator brush is making contact with a particular pole, a sample is taken of the amplitude of the information wave whose channel is associated with that pole. The sample is a pulse whose amplitude is identical to the amplitude of the wave during the interval of time when the sample was taken. Successive samples taken from the same channel form a train of pulses whose amplitudes vary directly as the amplitude of the sampled wave varies. The result is a train of nearly rectangular pulses that bear the characteristics of amplitude modulation.

2. Commutation - A Means of Selection

Commutation is a means of selecting the sample from one channel at a time. This is because the commutator brush makes contact with only one pole at a time. Because the rotor moves

DOCID: 3726612

at a uniform speed, there is always a fixed interval of time between successive contacts with the same pole. Therefore, if the rotor speed is known, it is possible to know exactly when the samples from a given channel will appear.

The spacing between pulses and the width of each pole are fixed. When the brush is moved off of one pole there is a short time interval before contact is made with the next pole due to the spacing between poles. As soon as contact is made with the succeeding pole a sample is taken representing the channel associated with it. Because all pole widths are identical, all samples are of the same duration, and because the pole spacings are identical, there is a fixed time interval between samples from adjacent channels. Thus, commutation provides sampling that is successive and recurrent, and therefore a means of individual channel selection is made possible.

3. Commutation - A Means of Multiplexing

As a means of sampling, commutation becomes a means of converting continuous data into discrete data. This conversion leaves time intervals between samples so that samples from other channels can be fitted in. Obviously, then, data must be discrete if it is to be multiplexed by time division.

Since commutation equally divides the time required for its cycle (a frame) among each of the channels to be sampled, each channel will be sampled at the same rate. Because of this, the samples from each channel will have the same PRF even though no two channels will be sampled at the same time. Identical PRF is a requirement for time division multiplexing so that pulses from each channel can be combined without any overlapping of samples. Commutation, then, provides a means for time division.

Although the commutator has many inputs, these, when sampled, are combined by the commutator into a single output. In this manner, a composite pulse train consisting of PAM samples from the several channels is synthesized. This pulse train synthesized by the commutator is a time multiplex signal that performs the same function in time division multiplexing as a baseband performs in frequency division multiplexing.

Commutation has the distinct advantage of combining sampling,

DOCID: 3726612

pulse amplitude modulation, and time division into a single operation, thus eliminating the need for much extra equipment (Figure 78).

4. Sub-commutation

The frequency response to information is much less for a time multiplex system than for a frequency multiplex or sub-carrier system. This generally presents no real problem in telemetry because about 90% of the information that is telemetered is quasistatic. In fact, some telemetry data varies at such slow rates that it is almost a waste of channel space in the system to sample it as frequently as once for each frame (commutator revolution). If there are several sources producing extremely slow information, the several information waves can be sampled and multiplexed onto a single commutator channel. This is done by means of a second commutator called a sub-commutator which samples such data less than once for each frame of the main commutator. The slowly varying signals are connected to the poles of the sub-commutator (thus they are called sub-commutator channels). These waves are then sampled and multiplexed by the sub-commutator into a single pulse train which is fed to one of the poles of the main commutator. The sub-commutator runs at a slower frame rate than the main commutator but is synchronized with it in such a way that all the sub-commutator channels are sampled in sequence, one for each frame of the main commutator.

Whenever there are a number of slowly varying inputs and it is necessary to accommodate more quasistatic channels, the desirable solution is to sub-commutate. The overall advantage may be summarized: slow information rates permit slow sampling rates, which, in turn, increase the number of quasistatic channels that can be accommodated if sub-commutation is employed.

Sub-commutation introduces sub-frames which naturally require the presence of sub-frame markers. (The existence of sub-frame markers, however, does not in itself prove that a system is sub-commutating.) These sub-frame markers must have some characteristic that distinguishes them from both the frame markers and from the information pulses. Like frame markers, the existence of sub-frame markers subtracts from the total number of channels that would be available to carry information.

DOCID: 3726612

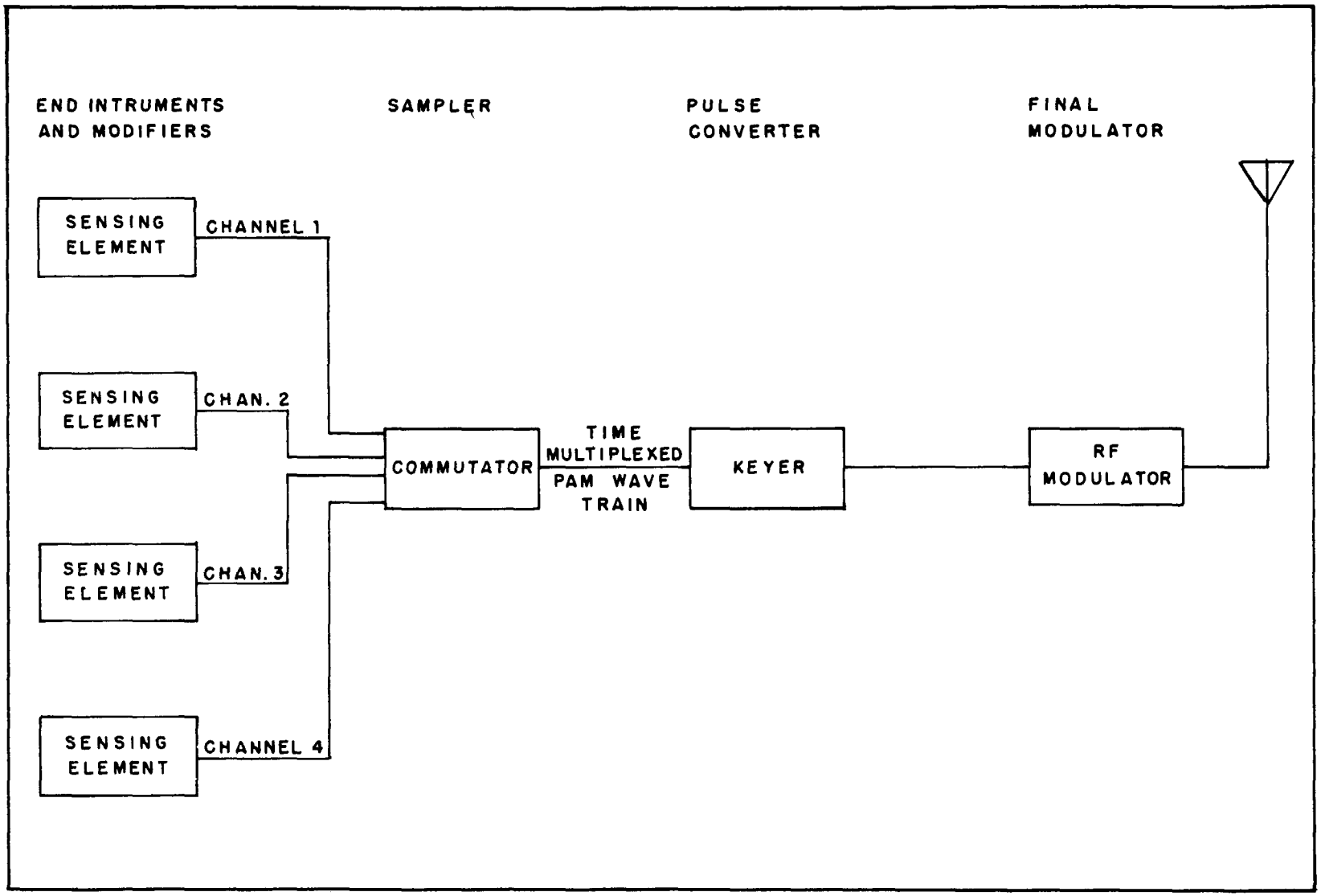


FIGURE 78. A REPRESENTATIVE TIME MULTIPLEXED TELEMETRY SYSTEM.

DOCID: 3726612

5. Strapping

The frequency response of a time sharing system can be increased for a given channel or channels, when necessary, by increasing the number of samples per frame for the information requiring the higher sampling rate. The modification is a simple one and can be accomplished without disturbing the rate at which the other channels are sampled. To do this, it is only necessary to cross connect the commutator pole associated with the channel in question to one or more additional unused poles equally spaced around the commutator. Poles strapped (cross-connected) in this manner must always be so spaced that there will be equal time intervals between each sample of the channel. Strapping increases the sampling rate to correspond to the higher frequency information but it is at the expense of decreasing the number of quasistatic information channels.

B. Time Division in U. S. Missile Telemetry

The type of time division multiplex system that has been standardized for missile telemetry in the United States uses pulse duration modulation. The PDM wave trains are developed from PAM wave trains which result from sampling the information by commutation methods. Time division systems are especially applicable to multiplexing of quasistatic data as such systems can handle many channels of information but are limited in their frequency response to slowly varying information.

1. Standard Commutation Parameters

Certain standards have been set up for commutation as applied to telemetry time multiplexing. The purpose of these standards is to establish a single fixed PRF for the commutator outputs while providing for sufficient flexibility to allow for several different systems. The standard also makes possible a simple method whereby these several systems can be identified.

(a) Number of Poles. Obviously, since the poles in a commutator are fixed in position, and since the number of poles in any given commutator cannot change, the brush contact on the rotor will make and break electrical connection with the same poles, in the same sequence for each frame or revolution. As every pole represents one sample, the number of poles establishes the number of samples that occur during

DOCID: 3726612

each revolution. For the four standard systems there are, respectively, 30, 45, 60, and 90 samples (poles) per frame.

(b) Commutator Speed. The commutator speed is the rate at which the rotor of the commutator turns. At first it might seem convenient to measure the commutator speed in terms of the number of revolutions the rotor makes each second. Although there is nothing wrong with this method of expression, the commutator speed can be thought of in slightly different terms that are more descriptive of time multiplex operation. A frame consists of one complete sequence of samples from the commutator inputs and includes the frame markers and calibration pulses. In practice in telemetry time division there is one complete frame for each revolution of the commutator rotor. Because of this one-to-one frame per revolution relationship, it is both more practical and more suggestive to express commutator speed in terms of the number of frames completed each second. It is for this reason, also, that commutator speed is frequently called the frame rate. The commutator speeds for the four standard systems are 30, 20, 15 and 10 frames/second, respectively.

(c) Commutation Rate. Commutation rate is the total number of samples taken by the commutator in a second of time. Two factors determine what the commutation rate will be. The first factor is the number of commutator poles (i.e., the total number of samples taken in each frame) and the second factor is the commutator speed (i.e., how many frames are completed by the commutator in each second). The product of these two factors is the commutation rate. In missile telemetry systems, there is a standard rate amounting to 900 samples per second. For example, for the 30 pole system which has a rotor speed of 30 frames per second, the commutation rate is,

$$\text{Poles X Rotor Speed} = \frac{30 \text{ samples}}{\cancel{\text{frame}}} \times \frac{30 \cancel{\text{frames}}}{\text{second}} = \frac{900 \text{ samples}}{\text{second}}$$

For the other three standard systems, computation of the commutation rate will result in the same value.

(d) Identification Method. Growing out of the standard on commutation rate (PRF of the commutator output) is a

DOCID: 3726612

method of identification of the commutation system. The thirty-pole commutator can have a rotor speed of only 30 frames per second. It is therefore called a thirty-by-thirty system (written 30 x 30). The forty-five-pole system can have a rotor speed of only 20 frames per second and is therefore called a 45 x 20 system. Similarly, there is a 60 x 15 system and a 90 x 10 system. In each instance the number of poles or samples per frame is given first and the rotor speed or frame rate is given second.

(e) Information Channels vs Number of Poles. As a note of caution, it must not be assumed that just because a 45 x 20 system has 45 poles or a 90 x 10 system has 90 poles, that there are an equal number of information channels. In these systems, the equivalent of two samples is used for frame markers. This reduces the possible number of information channels by two. Hence the maximum number of information channels that the 30 x 30, 45 x 20, 60 x 15, and 90 x 10 systems can carry are 28, 43, 58, and 88, respectively.

In order to increase the accuracy of the information it is necessary to electronically compare the telemetered data against a scale factor. This requires zero and full-scale calibration pulses to be sent along with the multiplexed information. These pulses occupy channel space and so further reduce the number of channels of quasistatic information that can be telemetered.

The number of quasistatic channels may still be less than the number of available commutator poles. This is determined solely by the requirements of the user. The commutation rate and the PRF of the commutator output, however, remain unchanged at 900 samples/second.

(f) Channel Sampling Rate vs Information Bandwidth. When the identity of a commutation system is known the sampling rate for any given channel is known automatically. Assuming that no poles are strapped together, each channel will be sampled only once for each frame. Continuing this assumption, the frame rate is also the number of samples per second per channel, or in other words, the quantity of frames per second will also be the quantity of samples per second for any given channel. The sampling rate per channel, then, for the 30 x 30,

DOCID: 3726612

45 x 20, 60 x 15, and 90 x 10 systems is, respectively, 30, 20, 15, and 10 samples per second per channel.

The number of samples per cycle of an information wave will establish the accuracy with which the information can be reproduced. The greater the number of samples taken of an information wave, the greater will be the accuracy when recovering the information from the samples. The desired accuracy will depend upon the needs of the user. Although the bare minimum sampling ratio from which information can be recovered is slightly more than two samples per cycle of the highest frequency component of the information wave, this does not provide sufficient accuracy for telemetry purposes. Even voice transmission requires the higher ratio of 2.4 samples per cycle of the highest modulating frequency. Telemetry signals require a high degree of accuracy, frequently to within .1%. To achieve this accuracy, the minimum sampling ratio must be set at 6 samples per cycle of the highest information frequency.

Together, the channel sampling rate (a system characteristic) and the number of samples per information cycle (a signal requirement) determine the maximum information rate (hence the information bandwidth) for any individual channel. The highest possible information rate for each channel is found by dividing the channel sampling rate by the required number of samples per information cycle. Thus, the highest information rate that the 30 x 30 system can accommodate is,

$$\frac{30 \text{ samples}}{\text{second}} \times \frac{\text{cycle}}{6 \text{ samples}} = \frac{5 \text{ cycles}}{\text{second}}$$

or an information bandwidth of dc to 5 cycles per second.

It should be noted here that the greater the number of channels that can be multiplexed, the smaller the allowable information bandwidth. This is because with a fixed commutation rate, an increase in the number of poles (potential channels) necessitates a decrease in the frame rate. This automatically decreases the channel sampling rate. To illustrate, compare the information bandwidth of the 45 x 20 and 60 x 15 systems with the 30 x 30 system previously computed.

DOCID: 3726612

Example 1, the 45 x 20 system.

$$\frac{20 \text{ samples}}{\text{second}} \times \frac{\text{cycles}}{6 \text{ samples}} = \frac{3.33 \text{ cycles,}}{\text{second}} \text{ or}$$

an information bandwidth of dc to 3.33 cps.

Example 2, the 60 x 15 system.

$$\frac{15 \text{ samples}}{\text{second}} \times \frac{\text{cycle}}{6 \text{ samples}} = \frac{2.5 \text{ cycles,}}{\text{second}} \text{ or}$$

an information bandwidth of dc to 2.5 cps.

2. Conversion of Pulse Modulation

The output of the commutator is a train of multiplexed PAM pulses. If another type of pulse modulation is desired such as PDM, PPM, or PCM, the train of PAM samples from the commutator must be converted. The device that performs the function of converting one type of pulse modulation into another is variously called a keyer, pulse converter, coder, or pulse modulator. A keyer performs its function by modifying the pulse characteristics in time, amplitude, or shape to produce the desired waveform.

In the current practice outlined in the telemetry standards, PDM pulses must comprise the time multiplexed wave train. In this instance the keyer converts the PAM wave into a PDM wave. The PDM pulses all have an equal, constant amplitude, but vary in duration in proportion to the amplitude of the respective PAM samples they replace. According to present standards, a sample having minimum amplitude is represented by a PDM pulse having a 90 microsecond duration and a sample having maximum amplitude is represented by a PDM pulse having a 700 microsecond duration. Between these two limiting values the relationship between the pulse duration and the amplitude of the information (samples) is linear. The PRF of the PDM pulse train is 900 pulses per second (the same as the PAM pulse train from the commutator) and corresponds to the commutation rate of 900 samples per second.

III. THE TIME DIVISION SIGNAL AS A MODULATING WAVE

The pulse carriers produced when sampling channels of information for time division multiplexing fulfill the same function as the subcarriers in frequency division. The several interleaved pulse subcarriers, when synthesized into a

DOCID: 3726612

composite pulse train by the commutator, are equivalent to the baseband used in frequency division multiplexing. The big difference between time multiplex and frequency multiplex is that in the time sharing systems, all pulse subcarriers have the same frequency (PRF) but the pulses representing the different channels are staggered in time, giving a time separation between samples of the respective inputs; whereas, in frequency sharing systems all the sinusoidal subcarriers are continuous in time but are separated in frequency with enough spacing that the sidebands from the adjacent channels will not overlap. Frequency division (not specifically in telemetry) makes use of any appropriate type of modulation for its subcarrier bands, and the modulation is done directly as A-M, P-M, F-M, SSB, and so on. In time division, on the other hand, the direct output of the commutator is always amplitude modulation (PAM) and then converted to the required form of pulse modulation by means of the keyer. In the same manner as a baseband is used to modulate some final sinusoidal carrier, the final form of the time multiplexed pulse train (whether a keyer is used or not) becomes a modulating wave for a sinusoidal carrier.

If there is to be a hybrid system of double multiplexing this sinusoidal carrier may be a subcarrier in a frequency division baseband, but in the case of pure time division, this will be the R-F carrier that is necessary for transmission through space. The latter case will be assumed in the following discussion.

A. The R-F Carrier

In order for the time multiplex wave train to be radiated, it is necessary for it to modulate a radio frequency carrier. The type of final modulation employed could be A-M, F-M, P-M, or some variation of these basic forms. In U. S. guided missile telemetry, the R-F carrier for final modulation by either a time division or a frequency division multiplex signal must be angle modulated. Either F-M or P-M is used.

In telemetry use, several bands are designated for these R-F carriers and their sidebands. They are 216 to 260 megacycles per second; 1435 to 1585 megacycles per second, with one-half of the band relegated to aeronautical flight testing and the other half relegated to missile and space vehicle testing; and 2200 to 2300 megacycles per second. Including drift and all other variations, the carrier operating within these band allocations must maintain a stability of .01% or better.

The separation between carriers depends upon the band or the part of the band that the carrier is in but the bandwidth of each channel around any particular carrier is 500 kilocycles. The maximum frequency deviation permitted (equivalent frequency deviation in the case of P-M) is ± 125 kilocycles per second, and the recommended minimum deviation is ± 75 kilocycles. When the maximum

DOCID: 3726612

frequency deviation is the one used, the guard band occurring at each end of the channel is 125 kilocycles per second.

B. Multiplex System Designations

In a previous chapter, it was described how frequency multiplex systems are identified by the letter symbols that designate the types of modulation employed. It is also customary to employ this same scheme of identification to time division systems. A letter group appears in the designation for each type of modulation used. Beginning with the input and working toward the modulation of the final carrier, the order in which these letter groups appear is indicative of the order in which the various modulations take place. To illustrate, some of the various time multiplex systems that have been employed in telemetry (not necessarily missile or space telemetry) are PCM/FM, PPM/AM, PAM/FM, PDM/FM and PDM/PM. The last two of this list, it should be recognized, are the methods currently standardized for use in missile and space telemetry. To give an indication of how these are read, PCM/FM is "pulse code modulation on frequency modulation." This means that the time division pulses constitute a pulse code, and that the multiplex wave train is used to frequency modulate an R-F carrier.

IV. BANDWIDTH REQUIREMENTS

Recording and telemetry almost invariably go hand in hand. This is especially true when the telemetry system multiplexes channels of simultaneous information. Recording provides a necessary convenience, making it possible to acquire data when it is available, and allowing for analysis and study at a later time. After reception and detection of an R-F signal carrying time multiplexed telemetry, there remains the pulse train. The pulse train may be recorded as is on a wide band recorder then at a later time demultiplexed and the individual channels demodulated. It is very important that the fidelity of the pulse train not be impaired by insufficient recorder bandwidth. This is sufficiently important as to warrant some consideration of what these bandwidth requirements are.

A. Pulse Bandwidth Requirements

By now the required bandwidth for a rectangular pulse or a train of rectangular pulses should be familiar ground to the reader. It should be recalled that since transient pulses have a continuous energy distribution instead of discrete spectral components, one cannot in any strict sense think of their bandwidth in terms of a certain number of harmonics but their bandwidth can be considered in terms of the harmonics of an equivalent squarewave. An equivalent squarewave is one whose pulse duration is equal to the duration of the transient rectangular pulse. It is safe to assume that any system that meets the minimum requirements for a squarewave of specified pulse duration will more than accommodate the minimum

DOCID: 3726612

requirement of any other rectangular pulse whose duration is not less than that of the squarewave.

A second thing to be remembered is that the bandwidth required for a train of rectangular pulses of unequal duration is that bandwidth required by the shortest element in the pulse train. It is this shortest element that requires the greatest bandwidth.

Still a third fact should be recalled, that although the least number of harmonics required to form a squarewave is the first (fundamental) and third, it is good practice to include up to the fourth harmonic even though the latter has no amplitude for a true squarewave. This bandwidth, in terms of the (sine X)/X curve includes everything from dc up to the second (sine X)/X dropout.

Now to apply these facts, consider a wave from a PDM system. The highest frequency in the PDM wave train will be established by the pulse having the greatest bandwidth. This pulse is, of course, the one with the shortest duration. For example, a PDM system whose pulse duration varies with modulation from 90 microseconds to 700 microseconds the required recording bandwidth is found by:

SHORTEST PULSE DURATION = W = 90 MICROSECONDS

PERIOD OF EQUIVALENT SQUAREWAVE = 2W = 180 MICROSECONDS

$$\text{FUNDAMENTAL FREQUENCY OF EQUIVALENT SQUAREWAVE} = \frac{\text{CYCLE}}{180 \text{ MICROSEC}} \times \frac{1 \times 10^6 \text{ MICROSECS}}{\text{SECOND}}$$

$$\frac{.00555 \times 10^6 \text{ CYCLES}}{\text{SECONDS}} = 5.55 \text{ KILOCYCLES/SECOND}$$

FOURTH HARMONIC FREQUENCY = 4 x 5.55 KC/S = 22.2 KC/S

Thus, the pulse train, in order to preserve the fidelity in all channels, requires a bandwidth of 22.2 kilocycles/second.

When bandwidth is determined in terms of the fourth harmonic of an equivalent squarewave, as in the example above, the computation can be shortened. The required bandwidth is double the reciprocal of the pulse duration, i.e., $2\left(\frac{\text{cycles}}{W}\right)$. The appropriate conversion factors must still be used to convert to cycles, kilocycles, or megacycles -- whichever is most desirable. Reworking the previous example by the shorter method:

$$BW = 2\left(\frac{\text{cycles}}{90 \text{ ~~microsec~~}}\right) \times \frac{1 \times 10^6 \text{ ~~microsec~~}}{\text{Second}} = 22.2 \text{ Kilocycles}$$

DOCID: 3726612

It should be understood that the bandwidth of each channel varies as the pulse duration changes due to modulation. It is also possible that some channels may not be modulated by a wave that will cause the minimum pulse duration. However, since many channels are on at the same time in a multiplex scheme such as this, it must be assumed that the maximum bandwidth will be required at all times.

B. R-F Bandwidth Requirements

The radio frequency bandwidth is always a function of the highest modulating frequency. This frequency symbolized by f_m has entered into all of the bandwidth equations when modulation was involved. These are:

$BW = 2 f_m$	Amplitude Modulation
$BW = 2 (f_d + f_m)$ or $2 f_m (M_f + 1)$	Frequency Modulation
$BW = f_m$	Single Sideband A-M
$BW = 2 f_m (M_p + 1)$	Phase Modulation

The R-F bandwidth requirements for either telemetry transmitters or receivers are computed in a straightforward manner using the equation that is appropriate for the variety of final modulation employed. The value of f_m to be substituted into the equation is the upper frequency limit of the recording bandwidth.

To illustrate, consider that the PDM wave in the example on recording bandwidth is to phase modulate the R-F carrier. Let it be assumed that the highest modulating frequency (22.2 Kc) in that recording band is able to cause a maximum phase deviation of 2 radians. Solving for the R-F bandwidth of the PDM/PM wave there obtains,

$$BW = 2 f_m (M_p + 1)$$
$$= 2 \times 22.2 \text{ Kc} \times (2 + 1)$$
$$= 133.2 \text{ Kc}$$

CONCLUSION

Time division multiplexing is one of the very useful and important applications of pulse theory and pulse techniques. The system is a method whereby separate pulse trains, each modulated by separate channels of information are interleaved so as to form a composite pulse train. From this fact, it follows that each successive pulse in the composite train represents a sample taken in a fixed time sequence from one of the information channels. Since the time occupied by the composite train is shared by all of the pulses from each of the channels of information, this is

DOCID: 3726612

called a system of time sharing or of time division multiplexing. This type of multiplexing is possible so long as the pulse repetition frequencies of the separate channels are kept in synchronism so as to maintain a constant progressive time difference between pulses from the different channels.

V. PRACTICAL EXERCISES

1. The characteristic of the information wave that is sampled is the

   (a) frequency
   (b) period
   (c) amplitude
   (d) phase

2. Pulse modulation in which pulses of varying widths occur is

   (a) PPM
   (b) PAM
   (c) PFM
   (d) PDM

3. Which of the following types of pulse modulation cannot be used for time division multiplexing?

   (a) PPM
   (b) PAM
   (c) PFM
   (d) PDM

4. Which of the following types of pulse modulation is not classed as pulse time modulation?

   (a) PPM
   (b) PAM
   (c) PFM
   (d) PDM

5. The minimum theoretical sampling ratio to reproduce the highest information frequency is:

   (a) 2 samples per cycle
   (b) 6 samples per cycle
   (c) 2.4 samples per cycle
   (d) 1 sample per 2 cycles

6. The desired accuracy for telemetry information is to within

   (a) 1%
   (b) .1%
   (c) .01%
   (d) 10%

7. Which of the following is not a particular type of pulse modulation?

   (a) PPM
   (b) PAM
   (c) PDM
   (d) PCM

8. Time division multiplexing is made possible by the pulse carriers from each channel having

   (a) a different PRF
   (b) a time interval between pulses
   (c) a different kind of modulation
   (d) varying pulse durations

9. Sampling, amplitude modulation, and time division can be accomplished simultaneously by

   (a) a keyer
   (b) a commutator
   (c) an end-instrument
   (d) a transmitter

A certain commutator produces 30 samples/frame. The commutator speed is 5 frames/second.

10. What is the commutation rate?

11. How much time is allotted for each sample?

DOCID: 3726612

12. If the pulse duration of each sample is exactly one-half of the period, what is the width of each sample?

13. If the highest information frequency to be sampled is 5 cycles per second and the desired number of samples per cycle is 6, what must the channel sampling rate be?

14. If the commutator speed is 5 frames/second, how many samples per frame must be taken of a channel requiring 30 samples/second channel sampling rate? (Strapping is assumed)

15. How would the poles in problem 14 have to be strapped in order to provide the proper sampling rate for the channel in question?

16. The commutator speed in revolutions per second is generally equal to

    (a) the number of channels per second
    (b) the channel sampling rate in samples/second per channel
    (c) the PRF in pulses per second
    (d) the information sampling ratio in samples per cycle

17. What is the principal advantage of PPM over PDM?

18. What is the advantage of PDM over PAM?

19. Why would the output of a commutator be applied to a "keyer"?

20. Why isn't PCM a distinct type of pulse modulation?

DOCID: 3726612

APPENDIX

For the convenience of users of this text, Part II of IRIG Document No. 103-56, Telemetry Standards for Guided Missiles (with revisions and addendum) has been included below. This part deals with the PDM/FM and PDM/PM Standards.

PART II PDM/FM OR PDM/PM STANDARD

1. GENERAL

The pulse duration modulation (PDM) systems are intended for use where a strictly time division multiplex system can meet the bulk of the telemetry requirements of a given application. A relatively large number of information channels can be accommodated but a relatively low frequency response capability in comparison with the subcarrier channels of the FM/FM system. Although the specifications to be given are based on previous experience with mechanical commutators, nothing in this standard is intended to prevent the use of electrical commutation methods.

2. PULSE DURATION MODULATION SPECIFICATIONS

The following are the specifications for the pulse duration modulated signal:

Number of samples per frame**	30	45	60	90
Frame rate (frames/sec)	30	20	15	10
Commutation rate (samples/sec)***	900	900	900	900

The information being transmitted in each channel shall determine the duration of the corresponding pulses. The relation between information magnitude and pulse duration should, in general, be linear.

Minimum pulse duration (zero level information)	90 ± 30 microseconds
Maximum pulse duration (maximum level information)	700 ± 50
Pulse rise and decay time (measured between 10% and 90% levels)	10 to 20 microseconds (constant to 3 microseconds for a given transmitting set)

The time interval between the leading edges of successive pulses within a frame shall be uniform from interval to interval and shall be constant within plus or minus 25 microseconds. This time interval shall have a nominal period equal to one

divided by the total sampling rate.

*These systems have sometimes been designated as PWM/FM or PWM/PM.

**The number of samples per frame available to carry information is two less than the number indicated because the equivalent of two samples is used in generating the frame synchronizing pulse.

***Commutation rate is equal to frame rate multiplied by the number of samples per frame.

The commutator speed (or frame rate) shall not vary more than plus 5% to minus 15% from nominal.

Frame synchronization of the receiving station shall be provided for by leaving a longer than normal time gap in the train of pulses transmitted. This gap shall be the same as that normally occupied by two successive data channel pulses. A representative of the pulse train wave form is shown in Figure 79.

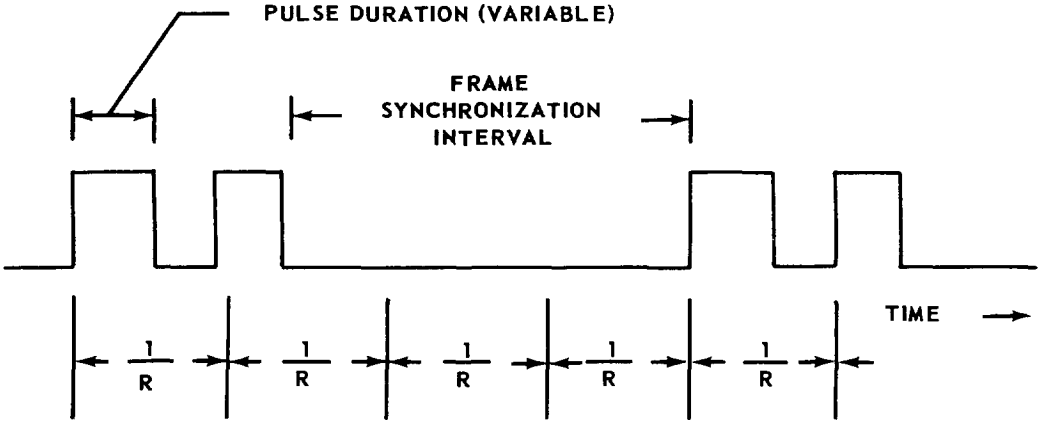


FIGURE 79

3. Radio Frequencies

The specifications concerning the radio frequency carrier to be used shall conform to IRIG Recommendation No. 101-59, "Telemetry Frequency Utilization Parameters and Criteria." (See Paragraph 5, FM/FM Standards.)

DOCID: 3726612

ANNEXES

ANNEX A - MATHEMATICS FOR MODULATION

ANNEX B - OSCILLOSCOPE PRINCIPLES

DOCID: 3726612

ANNEX A

MATHEMATICS FOR MODULATION

The study of mathematics can be compared to the study of a language. In fact, mathematics is a language--the language of numbers and size. When studying English, one must master the rules of grammar in order to completely understand the language. Likewise in mathematics there are many rules, concepts, definitions, terms and words that must be learned in order to completely understand the subject.

The subject matter that follows requires a bare minimum of mathematics. With the exception of dimensional analysis and powers of ten, no mathematical operations are used which are not included in the usual grade school and high school curricula. The intent and purpose of this section of the signal analysis course is to provide a review of the mathematical principles and operations that will find application in this course and which have usually been forgotten by the student who has been out of high school for several years. Inasmuch as addition and subtraction of fractions has found little or no application in this text, a review of these two operations along with common fractions and the associated finding of a least common denominator have been omitted. The information included here is developed under the topics of common fractions, decimal fractions, positive and negative numbers, powers of 10, and dimensional analysis.

I. COMMON FRACTIONS

A. The Basic Fraction

1. There are three ways in which we may indicate that one quantity is to be divided by another:

(a) $n\overline{)N}$

(b) $N \div n$

(c) $\frac{N}{n}$

We may add a bit of realism to the above expressions by assigning values to the general numbers n and N. For example, let $n = 3$ and $N = 4$. We then have:

DOCID: 3726612

(a) $3\overline{)4}$

(b) $4 \div 3$

(c) $\frac{4}{3}$

It will be noticed that form (c) is that of a fraction, but in all three cases the desired operation is the same and the answer is the same, $1\frac{1}{3}$.

2. When it is definitely intended that the operation of division be performed the (a) and (b) forms are generally used,

$3\overline{)4} = 1\frac{1}{3}$ or $4 \div 3 = 1\frac{1}{3}$

3. When it may be satisfactory to leave the division unperformed, it is written in form (c) as a fraction $\frac{4}{3}$. This form indicates that a division is intended. The fraction bar, then, is equivalent to a division sign.

4. When the divisor is greater than the dividend, the indicated operation may be properly left in fractional form unless, for a particular case, further solution is desirable. For example, let n = 2 and N = 1. We have then, (a) $2\overline{)1}$, (b) $1 \div 2$, and (c) $\frac{1}{2}$. Note that (a) and (b) both require a solution but form (c) may be considered as a solution in itself.

5. When a division is written in fractional form, the dividend is called the numerator, and the divisor is called the denominator. The numerator or dividend tells us how many parts or items there are while the denominator or divisor tells us the number of equal parts into which these are divided.

6. Denominator means name, and it is the denominator that provides the name for a given fraction. For example, if the denominator indicates that the item or items are to be divided into 3 equal parts, the parts are called thirds, or if divided into 12 equal parts, twelfths, etc. The numerator gives us the number of 3rds, 12ths, etc., that there are.

7. From the preceding statements, then, we can conclude that $\frac{2}{3}$ for example, may be read, "two-thirds" or "two divided by three", either with equal correctness provided the context will allow it. In either case the meaning is the same.

DOCID: 3726612

8. Kinds of Common Fractions

Common fractions fall into two categories as determined by the relative sizes of the numerator and denominator. These are proper fractions and improper fractions.

(a) Proper Fractions. Any common fraction whose numerator is smaller than the denominator is called a proper fraction. For example:

$$\frac{2}{3}, \frac{7}{16}, \frac{23}{25}, \frac{5}{12}, \frac{7}{8}, \frac{4}{5}$$

(b) Improper Fractions. Any common fraction whose numerator is equal to or greater than its denominator is an improper fraction. For example:

$$\frac{3}{3}, \frac{4}{4}, \frac{3}{2}, \frac{4}{3}, \frac{32}{5}, \frac{9}{8}$$

9. Properties of Fractions

If both the numerator and denominator of any common fraction are divided by the same number except zero the value of the fraction is unchanged. Fractions that are equal to each other but in different terms are called equivalent fractions.

$$\frac{3}{8} = \frac{3 \times 8}{8 \times 8} = \frac{24}{64}$$

$$\frac{24}{64} = \frac{24 \div 8}{64 \div 8} = \frac{3}{8}$$

B. Reduction of Fractions

Unless otherwise indicated, common fractions occurring in an answer should always be left in lowest terms and as a proper fraction (or whole number and proper fraction). The process involved in accomplishing this is called reduction.

1. Reduction of Proper Fractions

A fraction is in its lowest terms when there is no number that will evenly divide both the numerator and denominator. Lowest terms of a

DOCID: 3726612

fraction are found by dividing them both evenly. For example: reduce $\frac{24}{64}$ to lowest terms. The largest number that will divide both numerator and denominator evenly is 8.

$$\frac{24 \div 8}{64 \div 8} = \frac{3}{8}$$

Thus $\frac{24}{64}$ is reduced to an equivalent fraction having lowest terms. Another example, reduce $\frac{50}{100}$. The largest number that will divide both numerator and denominator evenly is 50.

$$\frac{50 \div 50}{100 \div 50} = \frac{1}{2}$$

2. Reduction of Improper Fractions

A whole number is called an integer. For example, 1, 4, 11, 5000 are integers. An integer and a fraction used together are called a "mixed number", e.g., $1\frac{1}{2}$, $6\frac{7}{8}$, $4\frac{2}{3}$ are mixed numbers. When an improper fraction is reduced to lowest terms there will result an integer, or an integer and a fraction (mixed number).

3. If an improper fraction is to be reduced, it should first be reduced to lowest terms and then the indicated division should be performed. If there is no remainder, the improper fraction has reduced to an integer. If, on the other hand, there is a remainder, the quotient will be a mixed number consisting of an integer and a fraction whose numerator is the remainder.

Example 1. Reduce the improper fraction $\frac{32}{8}$ to lowest terms and proper form.

$$\frac{32 \div 8}{8 \div 8} = \frac{4}{1} \qquad \text{(reduced to lowest terms)}$$

$$\frac{4}{1} = 4 \qquad \text{(reduced to proper form)}$$

DOCID: 3726612

Example 2. Reduce $\frac{34}{8}$ to lowest terms and proper form.

$$\frac{34 \div 2}{8 \div 2} = \frac{17}{4} \qquad \text{(reduced to lowest terms)}$$

$$\frac{17}{4} = 4\frac{1}{4} \qquad \text{(reduced to proper form)}$$

Dividing or multiplying any number by 1 will not change its value. It should be obvious that any fraction whose numerator and denominator are equal, such as $\frac{4}{4}, \frac{3}{3}, \frac{25}{25}$, etc., is equal to one. Therefore, multiplying or dividing any number, even a fraction, by a fraction whose numerator and denominator are equal is the same as multiplying or dividing by one and hence does not change the value.

The above principle is used in converting both integers and mixed numbers into improper fractions. Any whole number is a fraction whose denominator is 1, for example, $4 = \frac{4}{1}$, $5 = \frac{5}{1}$, $100 = \frac{100}{1}$, etc., but for convenience in everyday usage, we usually omit this denominator. Now let us reduce 7 to fourth's.

(a) Write the integer in fractional form: $7 = \frac{7}{1}$

(b) Multiply by $\frac{4}{4}$: $\frac{7}{1} \times \frac{4}{4} = \frac{28}{4}$

(c) Thus, 7 is equal to: $\frac{28}{4}$

Another example: reduce 8 to halves.

(a) Write the integer in fractional form: $8 = \frac{8}{1}$

(b) Multiply by $\frac{2}{2}$: $\frac{8}{1} \times \frac{2}{2} = \frac{16}{2}$

(c) Thus, 8 is equal to: $\frac{16}{2}$

A mixed number may be converted to an improper fraction just as easily as the foregoing examples. To illustrate, let us reduce $2\frac{7}{8}$ to an improper fraction. Notice that since part of a mixed number is already in fractional form, we have only to reduce the integer to a fraction. Inasmuch as the fractional part is in eighths we will want to reduce the integer to eighths also so they can be added together.

$$\frac{2}{1} \times \frac{8}{8} = \frac{16}{8}$$

DOCID: 3726612

Now we add the $\frac{16}{8}$ to the $\frac{7}{8}$ and find that $2\frac{7}{8}$, a mixed number, is equal to $\frac{23}{8}$, an improper fraction.

C. Operations with Common Fractions

The arithmetic operations are addition, subtraction, multiplication and division. Nothing will be said of addition and subtraction of common fractions in this review other than the statement that each fraction to be added or subtracted must have a denominator that is common to all and the numerators are then added or subtracted as required.

1. Multiplication of Common Fractions

(a) To multiply a fraction by a fraction, the numerators are multiplied together to obtain the new numerator and the denominators are multiplied together to obtain the new denominator.

Example 1: $$\frac{3}{7} \times \frac{5}{8} = \frac{3 \times 5}{7 \times 8} = \frac{15}{56}$$

Example 2: $$\frac{4}{5} \times \frac{7}{11} = \frac{4 \times 7}{5 \times 11} = \frac{28}{55}$$

(b) A process known as cancellation performed before the multiplication takes place will make it unnecessary to reduce the product after performing the operation of multiplication and will make the actual multiplication easier.

Example : $$\frac{4}{5} \times \frac{7}{16}$$

Solving without cancelling:

$$\frac{4}{5} \times \frac{7}{16} = \frac{4 \times 7}{5 \times 16} = \frac{28}{80}$$

but the answer is not in lowest terms so we must reduce $\frac{28}{80}$.

$$\frac{28 \div 4}{80 \div 4} = \frac{7}{20}$$

DOCID: 3726612

Using cancellation this same problem becomes:

$$\frac{\cancel{4}^{1}}{5} \times \frac{7}{\cancel{16}_{4}} = \frac{1 \times 7}{5 \times 4} = \frac{7}{20}$$

Here the answer is already in lowest terms. Notice that the cancellation was done by simply dividing one of the numerators and one of the denominators by some number that will divide both of them evenly. The process of cancellation can be performed on more than one numerator-denominator pair if a number can be found that will divide them evenly.

Example: $\frac{4}{5} \times \frac{15}{32}$

Four will reduce the numerator 4 and the denominator 32. Five will reduce the numerator 15 and the denominator 5. Cancelling we find:

$$\frac{\cancel{4}^{1}}{\cancel{5}_{1}} \times \frac{\cancel{15}^{3}}{\cancel{32}_{8}} = \frac{1 \times 3}{1 \times 8} = \frac{3}{8}$$

(c) When multiplying a whole number by a fraction the whole number should be changed to a fraction with a denominator of one then the multiplication performed as with any other fraction.

$$40 \times \frac{7}{8} = \frac{\cancel{40}^{5}}{1} \times \frac{7}{\cancel{8}_{1}} = \frac{5 \times 7}{1 \times 1} = \frac{35}{1} = 35$$

(d) When multiplying a fraction by a mixed number, the mixed number should be reduced to an improper fraction by the method already described and the multiplication performed as with any other fraction.

$$2\frac{2}{3} \times \frac{1}{6} = \frac{\cancel{8}^{4}}{3} \times \frac{1}{\cancel{6}_{3}} = \frac{4 \times 1}{3 \times 3} = \frac{4}{9}$$

2. Division of Common Fractions

(a) Inversions and Reciprocals. A term often encountered in mathematics is "reciprocal." The reciprocal of a number is defined as 1 divided by that number. Expressed as a fraction, which we have learned is an <u>indicated</u> division, the reciprocal of N is $\frac{1}{N}$, the reciprocal of 2 is $\frac{1}{2}$, the reciprocal of 50 is $\frac{1}{50}$, etc.

DOCID: 3726612

It has already been explained that a whole number is a fraction with a denominator of 1. Applying this to the three examples just used above, we obtain, the reciprocal of $\frac{N}{1} = \frac{1}{N}$, the reciprocal of $\frac{2}{1} = \frac{1}{2}$, and the reciprocal of $\frac{50}{1} = \frac{1}{50}$. It should, then, be obvious that the reciprocal of an integer is simply found by expressing the whole number as a fraction and then inverting (turning upside down) it.

To divide with fractions we must invert the divisor and multiply, that is, we multiply by the reciprocal of the divisor. This may be illustrated by

$$\frac{2}{3} \div 2 = \frac{2}{3} \div \frac{2}{1}$$

Inverting the divisor and multiplying

$$\frac{\cancel{2}^{1}}{3} \times \frac{1}{\cancel{2}_{1}} = \frac{1}{3}$$

Now suppose we were to find the reciprocal of a fraction. How would we do it? The definition says that the reciprocal of a number is 1 divided by that number. This definition applies even when the desired reciprocal is that of a fraction.

$$1 \div \frac{1}{4} = \text{reciprocal of } \frac{1}{4}$$

The rule says "to divide with fractions, invert the divisor and multiply." This becomes, then

$$1 \times \frac{4}{1} \text{ or } \frac{1}{1} \times \frac{4}{1} = \frac{4}{1} = 4$$

Observe that finding the reciprocal of a fraction is done the same way as finding the reciprocal of whole numbers, that is, we invert the fractional form. Thus the reciprocal of $4 = \frac{1}{4}$, and the reciprocal of $\frac{1}{4} = 4$.

(b) Dividing Fractions by Fractions. In explaining what reciprocals are and how they are obtained, the method of dividing with fractions has actually been illustrated already. As in the case of ordinary multiplication with fractions, integers and mixed numbers should be reduced to improper fractions before performing the operation. Cancellation and reduction should be used wherever applicable just as in multiplication. The five examples that follow illustrate the possibilities that can occur when performing division with fractions.

DOCID: 3726612

Example 1: A fraction divided by a fraction.

$$\frac{5}{9} \div \frac{3}{4} = \frac{5}{9} \times \frac{4}{3} = \frac{20}{27}$$

Example 2: A fraction divided by an integer.

$$\frac{3}{5} \div 6 = \frac{3}{5} \div \frac{6}{1} = \frac{\cancel{3}^{1}}{5} \times \frac{1}{\cancel{6}_{2}} = \frac{1}{10}$$

Example 3: An integer divided by a fraction.

$$10 \div \frac{1}{4} = \frac{10}{1} \times \frac{4}{1} = \frac{40}{1} = 40$$

Example 4: A fraction divided by a mixed number.

$$\frac{7}{8} \div 1\frac{7}{16} = \frac{7}{8} \div \frac{23}{16} = \frac{7}{\cancel{8}_{1}} \times \frac{\cancel{16}^{2}}{23} = \frac{14}{23}$$

Example 5: A mixed number divided by a fraction.

$$4\frac{2}{3} \div \frac{5}{7} = \frac{14}{3} \times \frac{7}{5} = \frac{98}{15} = 6\frac{8}{15}$$

EXERCISE I

REDUCE TO LOWEST TERMS:

1. $\frac{32}{72} =$
2. $\frac{100}{300} =$
3. $\frac{24}{36} =$
4. $\frac{15}{35} =$
5. $\frac{126}{144} =$
6. $\frac{96}{224} =$

REDUCE TO MIXED NUMBERS:

7. $\frac{47}{16} =$
8. $\frac{33}{12} =$
9. $\frac{9}{8} =$
10. $\frac{3}{2} =$
11. $\frac{9}{4} =$
12. $\frac{256}{9} =$

DOCID: 3726612

REDUCE TO IMPROPER FRACTIONS:

13. $5\frac{7}{16}$ =

14. $16\frac{2}{3}$ =

15. $1\frac{5}{32}$ =

16. $11\frac{2}{25}$ =

17. $4\frac{125}{132}$ =

18. $6\frac{7}{8}$ =

ADD AND REDUCE:

19. $\frac{2}{3}$ + $\frac{5}{3}$ = _____

20. $\frac{7}{16}$ + $\frac{9}{16}$ = $\frac{11}{16}$ _____

21. $\frac{1}{23}$ + $\frac{4}{23}$ = $\frac{18}{23}$ _____

SUBTRACT AND REDUCE:

22. $\frac{18}{11}$ − $\frac{7}{11}$ = _____

23. $\frac{17}{35}$ − $\frac{10}{35}$ = _____

24. $\frac{20}{21}$ − $\frac{13}{21}$ = _____

MULTIPLY OR DIVIDE AS INDICATED:

25. $\frac{2}{5} \times \frac{7}{16}$ =

26. $3 \div \frac{2}{3}$ =

27. $6\frac{3}{13} \div 8\frac{2}{5}$ =

28. $15 \times \frac{2}{3}$ =

29. $\frac{22}{7} \div 6\frac{2}{7}$ =

30. $\frac{\frac{6}{15}}{\frac{4}{5}}$ =

DOCID: 3726612

II. DECIMAL FRACTIONS

A. The Basic Fraction

Fractions are of two kinds--the common fraction which may be either proper or improper and is an indicated division, and the decimal fraction. Just as common fractions are frequently referred to simply as "fractions," so decimal fractions are frequently called "decimals."

Decimal fractions are described as fractions that have some integral power of 10 such as 10, 100, 1000, etc., for denominators. In writing them, however, it is convenient to omit the denominator and in its place write a "point" (.) in the numerator called a decimal point. The decimal point is moved to the left in the numerator by as many places as there are zeros in the denominator. The decimal point in an integer is located to the right of the number but is usually omitted if there is no decimal fraction to follow. For example 2., 6., and 70., are normally written 2, 6, and 70. The fraction $\frac{32}{100}$ may easily be written in decimal form by writing down the numerator (32) as a whole number with the decimal point all the way to the right (32.), count the number of zeros there are in the denominator (there are two) and move the decimal point to the left by that many places (.32.). The common fraction $\frac{32}{100}$ then becomes .32, a decimal fraction.

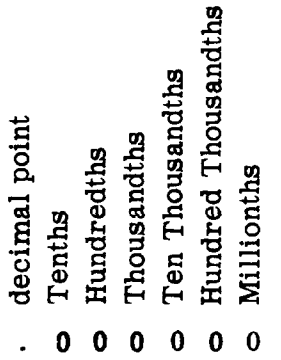
Suppose our example had been $\frac{32}{1000}$. Here the same procedure is followed as before but in order to move the decimal point the required number of places (three in this instance) a zero must be added to the left of the numerator and the result is .032. A rule then would be, if there are more zeros in the denominator than there are digits in the numerator, zeros must be added to the left of the numerator until there are as many places in the numerator as there are zeros in the denominator.

A third example might be the improper fraction $\frac{32}{10}$. This by the same procedure becomes 3.2, an integer 3, and .2 a decimal fraction.

DOCID: 3726612

B. Naming the Decimal Fraction

The position occupied by a digit is called a "place." Each decimal place has a name which is determined by its position to the right of the decimal point as illustrated:



A decimal fraction is named for the place occupied by the last digit of the fraction, thus .47 has its last digit in the hundredths place, the fraction is forty-seven hundredths, and .0577 has its last digit in the ten-thousandths place, hence this fraction is five hundred seventy-seven ten thousandths.

A decimal fraction is read in much the same manner as a common fraction. Thus .3, .73, and .47 are read three tenths, seventy-three hundredths, and forty-seven hundredths respectively just as $\frac{3}{10}$, $\frac{73}{100}$, and $\frac{47}{100}$ would be.

C. Annexing and Removing Zeros

The operations with decimals are frequently simplified by adding or removing zeros. Caution must be used in this process, however, or else the value of the decimal may be changed. Let's use the value 25.73 for our illustration.

1. Zeros may be added to the left of the whole number without changing the value:

$$0025.73 = 25.73$$

2. Zeros may be added to the right of the fractional part of a decimal quantity without changing the value:

$$25.7300 = 25.73$$

DOCID: 3726612

3. Zeros cannot be added to the right of the integer (but to the left of the decimal point) as this would affect the value:

$$2500.73 \neq 25.73$$

4. Zeros cannot be added to the left of the fractional part of a decimal quantity (but to the right of the decimal point) as this would affect the value:

$$25.0073 \neq 25.73$$

5. In addition to the above conditions, whenever it is possible to annex zeros, it is also possible to remove zeros without violation of the value.

D. Fraction to Decimal and Decimal to Fraction Conversion

It was stated earlier that a common fraction is an indicated division, and that a division so indicated is usually left in fractional form. When the indicated division is a proper fraction and the division is performed, the result (quotient) will invariably be a decimal fraction. As two common fractions equal to each other but in different terms are called equivalent fractions (for example $\frac{3}{4}$ and $\frac{6}{8}$) so a decimal fraction equal to a common fraction is called a decimal equivalent (for example .3 is the decimal equivalent of $\frac{3}{10}$).

1. Converting a Proper Common Fraction to a Decimal

(a) Rearrange the fraction in the form of a conventional division problem with the denominator as the divisor and the numerator as the dividend.

(b) Place a decimal point to the right of the dividend.

(c) Place a decimal point in the quotient directly above the decimal point in the dividend.

(d) Since the dividend is smaller than the divisor, zeros must be annexed to the right of the decimal point in the dividend in order to perform the division.

(e) Annex as many zeros to the dividend as are needed to obtain a decimal equivalent that is either without a remainder or to the desired accuracy.

DOCID: 3726612

The above five steps will be exemplified by converting $\frac{4}{25}$ to its decimal equivalent.

Step (a) Rearrange:

$$\frac{4}{25} \qquad 25\overline{)4}$$

Step (b) Place point in dividend:

$$25\overline{)4.}$$

Step (c) Place point in quotient:

$$25\overline{)4.}$$

Step (d) Annex a zero and divide:

$$\begin{array}{r} .1 \\ 25\overline{)4.0} \\ \underline{25} \\ 15 \end{array}$$

Step (e) Annex another zero and complete the division:

$$\begin{array}{r} .16 \\ 25\overline{)4.00} \\ \underline{25} \\ 150 \\ \underline{150} \\ 0 \end{array}$$

Thus, the decimal equivalent of $\frac{4}{25}$ is .16.

2. Converting a Decimal Fraction to a Common Fraction

A decimal fraction is easily converted into a common fraction by following this procedure.

(a) Count the number of places the decimal point would have to be moved to the right to make the complete decimal into a whole number.

DOCID: 3726612

(b) Remove the decimal point and give the remaining quantity a denominator consisting of 1 followed by as many zeros as there were decimal places in Step (a).

(c) Reduce the resultant common fraction to lowest terms.

Note how the illustration follows these steps:

Step (a) Two decimal places.

.85

Step (b) Conversion to common fraction.

$$\frac{85}{100}$$

Step (c) Reduction to lowest terms.

$$\frac{17}{20}$$

E. Operations with Decimals

1. Addition and Subtraction

Decimals may be added and subtracted in exactly the same manner as whole numbers provided the quantities are placed in columns with the decimal points all in line. If the addition or subtraction of decimals is attempted without keeping the points in line, erroneous results are sure to ensue. The right and wrong arrangements are shown for adding 2.563, 8.96, .001, 708. and 21.

WRONG	RIGHT
2.563	2.563
8.96	8.96
.001	.001
708.	708.
21.	21.

DOCID: 3726612

2. Multiplication with Decimals

Except for the placing of the decimal point, multiplication with decimals is no different from multiplication with ordinary whole numbers. To place the decimal point, there must be as many decimal places in the product as there are total decimal places in the factors being multiplied.

Example:

```
  3.875     3 places in the multiplicand
   43.5     1 place in the multiplier
 -------
  19375
 11625
15500
 -------
168.5625    4 places in the product
```

3. Division with Decimals

Long and short division with decimals is performed in exactly the same manner as with integers, the only difference occurring in the proper placement of the decimal point.

(a) When the divisor is a whole number, the division should be performed in the normal manner, with zeros annexed to the dividend as needed. The decimal point in the quotient is placed directly above the point in the dividend.

Example 1:

```
      3.54
   ------
54 | 191.16
     162
     ---
      291
      270
      ---
       216
       216
       ---
         0
```

Example 2:

```
      .00472
   ---------
162 | .76464
      648
      ----
      1166
      1134
      ----
        324
        324
        ---
          0
```

DOCID: 3726612

(b) When the divisor is either a decimal fraction or mixed number, the divisor must be cleared of decimal places by moving the point all the way to the right; then in order to preserve the value of the problem, the decimal point in the dividend must be moved to the right by an equal number of places. If there are not enough digits in the dividend to move the decimal point the required number of places, then zeros must be annexed to the right of the dividend (a principle already learned) so that the point can be moved as required. The point appearing in the answer must then be placed above the new decimal point in the dividend.

Example:

$.76\overline{)437}$

In order to move the decimal point in the dividend to the required place, two zeros must be annexed.

$.76\overline{)437.00}$

The point is moved as many places in the dividend as are needed to clear the divisor.

$$\begin{array}{r} 575. \\ 76\overline{)43700.} \\ 380 \\ \hline 570 \\ 532 \\ \hline 380 \\ 380 \\ \hline 0 \end{array}$$

EXERCISE II

ARRANGE IN COLUMNS AND ADD:

1. 43.1, 72.6, .009, .031, and 76.

DOCID: 3726612

2. 21, 40, .833, 60.007, 5.0.

3. 50, 72, 100.8, 65, .065.

ARRANGE IN COLUMNS AND SUBTRACT:

4. 2750. - 16.3

5. 103.5 - .0075

EXPRESS IN WORDS:

6. .0731

7. .00005

CONVERT TO DECIMAL FRACTIONS:

8. $\frac{7}{8}$

9. $\frac{2}{3}$

10. $\frac{1}{20}$

11. $\frac{1}{125}$

12. $\frac{4}{5}$

13. $\frac{5}{16}$

14. $\frac{27}{64}$

15. $\frac{22}{7}$

DOCID: 3726612

CONVERT TO COMMON FRACTIONS IN LOWEST TERMS:

16. .34375

17. .96875

18. .0001

19. .875

20. .625

OBSERVING DECIMAL POINT, MULTIPLY OR DIVIDE AS INDICATED:

21. .787
x 3.65

22. 4.32
x 7.15

23. .000015
x .036

24. .875 ÷ .125

25. .96875 ÷ .0001

26. .875 ÷ .625

27. .000707
x 9.847

28. .324 ÷ 45

29. 647
x .02

DOCID: 3726612

III. POSITIVE AND NEGATIVE NUMBERS

A positive number can be defined as any number greater than zero. Whenever such numbers are added, multiplied, or divided, the results are always positive. Likewise, when such a number is subtracted from a larger number the result will be positive.

What happens, though, when necessity requires us to subtract a number from a smaller one? Basic arithmetic provides no process for this operation. However, such instances do arise and it becomes necessary to deal with so-called negative numbers. A negative number can be considered as a number less than zero. Such numbers are always preceded by a minus sign. A positive number may, on the other hand, be preceded by a plus sign although any number which bears no sign is assumed to be positive.

A. Addition

Two trains leave Indianapolis, Indiana at 8:00 a.m. Two hours later train A is 60 miles and train B is 90 miles from Indianapolis. How far is B from A? Actually, the answer to this problem depends on whether the trains are traveling in the same direction or in some other direction.

Two possibilities are shown in the graphic illustrations:

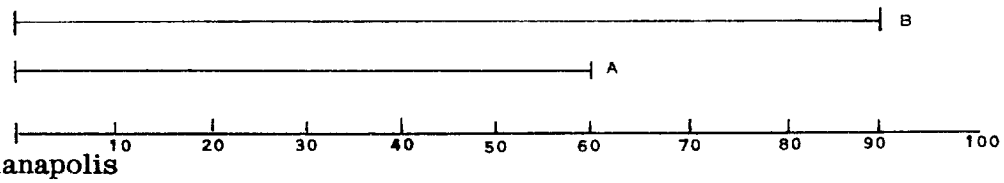


FIGURE 80a. SAME DIRECTION--A 30 MILES FROM B.

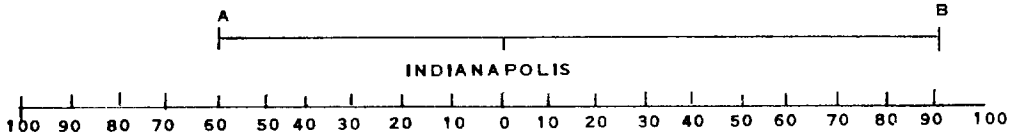


FIGURE 80b. OPPOSITE DIRECTION--A 150 MILES FROM B

With the graph in each case both length and direction have been shown. Figure 80a illustrates the problem assuming the trains travel in the same direction. Each unit represents 10 miles. Count the units from A to B. Your answer should be 30 miles. That is, train B is 30 miles from train A. Figure 80b illustrates the

DOCID: 3726612

problem assuming the trains travel in opposite directions. Train A has traveled 60 miles to the left or West of Indianapolis while train B has traveled 90 miles to the right or East of Indianapolis. Count the units from A to B with each unit again representing 10 miles. The distance from B to A is now 150 miles.

In a previous paragraph a so-called negative number was defined as being less than zero. The question probably arises: How can there be a number less than zero when zero is nothing? In the old school, the symbol for zero was called "naught" and naught means nothing. Present day teachers often correct their students and tell them not to call the symbol "naught" but to call it "zero." Few of these teachers give a reason for the change. As a matter of fact, few of them know a reason.

Contrary to popular beliefs the zero means something other than nothing. Actually, zero is a reference point. Temperature readings are taken with reference to a definite point on the thermometer, called the zero point. When the temperature reaches the zero point, it is said to be at the freezing point (centigrade thermometer). In some climates, though, the temperature frequently is much colder than freezing. These temperatures are referred to as "so many degrees below zero." Ten degrees above zero would simply be written as 10°. Ten degrees below zero, however, remembering that a so-called negative number is less than zero, would be written -10°

Engineers use as a zero or reference point the ground or the earth. This does not mean, however, that there are no potentials below the ground potential. There are times when the negative number may represent a larger absolute value (absolute value being the value of any number regardless of sign) than some positive number. The positive and negative signs simply denote reference from zero.

It has already been shown that two or more positive numbers can be added with no difficulty and that the sum will be positive. Also, when two or more negative numbers are added, the sum will be negative. In other words, to add 2 or more numbers with like signs, find the sum of their absolute values and prefix this sum with the common sign.

When a positive number is added to a negative number a different rule applies. To add a positive number to a negative number, find the difference of their absolute values and prefix to the result the sign of the one that has the greater absolute value.

DOCID: 3726612

When three or more algebraic numbers are to be added, differing in signs, find the sum of the positive numbers and then the sum of the negative numbers. Add these sums algebraically, using the rule for adding positive and negative numbers to obtain the total algebraic sum.

Examples: ADD

(a)	(b)	(c)	(d)
15	-15	15	-15
8	- 8	- 8	8
23	-23	+ 7	- 7

(e)	(f)	(g)	(h)
35	-236	-45	236
-38	780	-20	-780
24	-95	60	95
-15	45	55	- 45
6	494	50	-494

Notice that in examples (a), (b), (c) and (d) the same absolute values are used. However, the change of one or both signs causes a change in the sum. Likewise, examples (f) and (h) use the same absolute values but the signs were reversed. Thus, the absolute value of the sum is the same while the sign prefixing the absolute value indicates there is a vast difference between the two results.

PRACTICE PROBLEMS A:

ADD:	(a)	(b)	(c)	(d)	(e)
	10	-77	-56	19	12
	-19	29	-14	36	0
	30	-16	69	0	-43
	70	37	13	-101	29

B. Subtraction

Using the thermometer as an example once again, note that the change in the position of the mercury column illustrates the difference between two successive readings. If the first reading is 10° and the second is 6°, the change is the difference between 10° and 6°. The difference is found by counting from the second reading to the first. If the direction of counting is upward the difference is positive; if downward the difference is negative.

DOCID: 3726612

Subtraction of positive numbers can be considered or defined as the inverse of addition. That is, when a number is to be subtracted from a larger number, the question should be asked: "What number, when added to the smaller will result in the larger?" Example: Subtract + 6 from + 10. Or what number when added to + 6 will result in +10? Regardless of the way it is said (+10) - (+6)= + 4.

It is necessary that a distinction be made between the terms used in subtraction. This can best be done as follows:

Minuend	+ 987
Subtrahend	+ 789
Remainder	+ 198

Using the same absolute values as in the above example, the subtraction rule can be illustrated.

(a)	(b)	(c)	(d)
987	-987	987	-987
789	789	-789	-789
198	-1776	1776	-198

As shown in the examples, to subtract one number from another, change the sign of the subtrahend and add algebraically.

PRACTICE PROBLEMS B:

SUBTRACT:	(a)	(b)	(c)	(d)	(e)
	-62	-23	-42	47	-6
	-19	23	17	-23	-19

C. Multiplication

When multiplying positive and negative numbers the process is basically the same as if multiplying numbers that have no signs. Therefore, the rule for multiplication of positive and negative numbers concerns the signs. When two numbers having like signs are multiplied the answer (product) will be positive. When two numbers having unlike signs are multiplied the answer (product) will be negative. As in arithmetic when several numbers are to be multiplied, the product is taken of the first and second, then the result and third, etc.

DOCID: 3726612

Examples:

(a) (5) (3) = +15 (b) (-5) (-3) = +15

(c) (-5) (+3) = -15 (d) (+5) (-3) = -15

PRACTICE PROBLEMS C

MULTIPLY: (a) $\begin{array}{r} -15 \\ \underline{-\ 9} \end{array}$ (b) $\begin{array}{r} 72 \\ \underline{-21} \end{array}$ (c) $\begin{array}{r} 24 \\ \underline{-\ 9} \end{array}$ (d) $\begin{array}{r} -96 \\ \underline{-23} \end{array}$ (e) $\begin{array}{r} 36 \\ \underline{25} \end{array}$

D. Division

The process of dividing in algebra has the same meaning as dividing in arithmetic. For review purposes the following definitions are given:

1. Dividend - number, or quantity, that is to be divided.

2. Divisor - number by which a number, or quantity, is to be divided.

3. Quotient - result obtained by division.

That is, $\frac{\text{Dividend}}{\text{Divisor}} = \text{Quotient}$ or, $\text{Divisor} \overline{)\ \text{Dividend}}$ Quotient

In our discussion of subtraction of signed numbers, it was brought out that subtraction is the inverse of addition. Multiplication can be considered as a shortened form of addition. And now it will be shown that Division is the inverse of multiplication.

In multipication (+4) (+3) = +12 means that (+4) + (+4) + (+4) = +12. Also (+3) (+4) = +12 means (+3) + (+3) + (+3) + (+3) = +12. Since this is true then (+12) ÷ (+3) = +4 and (+12) ÷ (+4) = +3.

DOCID: 3726612

The rule for dividing signed numbers is the same as for multiplying, i.e., To divide positive and negative numbers,

1. if dividend and divisor have like signs, the quotient is positive.

2. if dividend and divisor have unlike signs, the quotient is negative.

Examples:

(a) $(+12) \div (+3) = +4$

(b) $(+12) \div (-3) = -4$

(c) $(-12) \div (+3) = -4$

(d) $(-12) \div (-3) = +4$

PRACTICE PROBLEMS D

DIVIDE: (a) $-108 \overline{)\ 36}$ (b) $-23 \overline{)\ 46}$

(c) $-12 \overline{)\ -96}$ (d) $-14 \overline{)\ 42}$

(e) $48 \overline{)\ -144}$

E. Summary of Positive and Negative Numbers

Positive and negative numbers are indicated by signs. A plus sign may or may not be used to signify that a number is positive because any number without a sign is considered positive. A negative number will always be preceded by a minus sign.

To add two or more numbers with like signs, find the sum of their absolute values and prefix the sum with the common sign. To add a positive number to a negative number, find the difference of their absolute values and prefix to the result the sign of the one that has the greater absolute value. When three or more algebraic numbers are to be added, differing in signs, find the sum of the positive numbers and then the sum of the negative numbers. Add these sums algebraically, using the rule for adding positive and negative numbers to obtain the total algebraic sum.

DOCID: 3726612

To subtract one signed number from another, change the sign of the subtrahend and add algebraically.

When two numbers having like signs are multiplied the answer will be positive. When two numbers having unlike signs are multiplied the answer will be negative.

The rule for division of positive and negative numbers says basically the same thing. That is, if dividend and divisor have like signs the quotient will be positive; and if dividend and divisor have unlike signs the quotient will be negative.

In working with signed numbers it is most important that the student remember what operation is being performed and always to recheck to insure the proper sign has been assigned the result.

EXERCISE III

ADD:

1.
```
  1459
- 287
    49
  3507
```

2.
```
-57.35
  2.87
 19.43
  5.09
```

3.
```
  507
 1014
  253
-2028
```

4.
```
  10495
- 3236
- 7059
    86.75
- 146.18
    30.15
```

5.
```
- 510
 -714
 -875
 -309
 -511
 -741
```

6.
```
 3478
 2679
 8276
 3001
 1905
 7790
```

SUBTRACT:

7. 9876 − 3456

8. −8052 − 4521

9. −8103 − 5790

10. −6208 − 3211

11. −2834 − 459

12. −8.753 − 2.4

13. 3.007 − .549

14. −7357 − 4530

15. 3450 − 2345

16. −6802 − −2680

17. 9579 − 3753

18. −4574 − −2077

MULTIPLY:

19. 5089 × 40

20. 10579 × −908

21. −834 × −59

22. −5790 × 904

23. 45302.75 × −.753

DIVIDE:

24. $-834\overline{)54210}$

25. $452\overline{)-2895}$

26. $-904\overline{)-2618888}$

27. $30\overline{)2850}$

28. $-17\overline{)-289}$

DOCID: 3726612

V. DIMENSIONAL ANALYSIS

Until now we have emphasized methods of obtaining correct numerical values in operations with common fractions, decimal fractions, signed numbers, and powers of ten. At this point the student should become acquainted with dimensional analysis. Dimensional analysis is a simplified method of solving problems by temporarily disregarding numerical values and considering only the dimensions; thereby eliminating the need for memorizing formulas. Solving an equation with dimensional analysis will be somewhat similar to the method employed in multiplying and dividing fractions: whereby cancellation may be performed if the same number is divided into the numerator and denominator. When applying dimensional analysis, cancellation of the numerators and denominators in the fractions set forth will be primarily applicable to the dimension of known facts rather than the numerical values.

A. Dimensions

Dimensions are measureable or observable characteristics of physical objects or abstract ideas. In this course the emphasis will be on bauds, cycles, and units of length and time. Dimensions are not always independent of each other. Frequently, they consist of a combination of other dimensions as in the case of velocity which combines distance and time (miles per hour). There are situations under which a dimension composed of two or more dimensions leaves the units of one of the dimensions unmentioned, e.g., cycles, kilocycles and megacycles are often used in place of cycles per second, kilocycles per second, and megacycles per second. This usage, although not quite correct, is acceptable as one of the shortcuts of the trade, the unmentioned term being understood by the user. In computations utilizing these terms the user must consider the otherwise unmentioned unit. As occasion and convenience may require, different units are sometimes used to express the same dimension as millisecond, microsecond, and second. But in this latter case, as the student will recognize, the change from one of the units to another requires only the shift of a decimal point or use of a different power of ten.

B. Expression of Terms in Dimensional Analysis

The type of answer most frequently sought for in dimensional analysis normally takes the form of "dimension per dimension" as illustrated by bauds per cycle, cycles per second, milliseconds per cycle, and milliseconds per baud.

These expressions when written in mathematical form appear as:

$$\frac{\text{bauds}}{\text{cycle}}, \quad \frac{\text{cycles}}{\text{second}}, \quad \frac{\text{milliseconds}}{\text{cycle}} \quad \text{and} \quad \frac{\text{milliseconds}}{\text{baud}}.$$

Observe that the word per is replaced by a fraction bar placed between dimensions. Per, in this usage, means divided by. The fractional form here has the same meaning as when used with pure numbers, namely, that there is an indicated division.

C. Dimensional Analysis Procedure

The procedure for solving problems through dimensional analysis may be included in five simple steps.

1. State to yourself the actual answer that is desired such as milliseconds per baud, cycles per second, bauds per cycle, etc.

2. Use mathematical form and write out the dimensions of the desired answer as stated but as though it were a fraction $\left(\frac{\text{milliseconds}}{\text{baud}}, \frac{\text{cycles}}{\text{second}}, \frac{\text{bauds}}{\text{cycle}}\right)$.

3. If known dimensions in the problem have units matching units of the desired answer, arrange them to coincide in position with the desired answer. This is illustrated by an example where $\frac{\text{Millisecond}}{\text{baud}}$ is the desired answer and the information available is 162 milliseconds per cycle and 7 bauds per cycle.

$$\frac{162 \text{ milliseconds}}{1 \text{ cycle}} \times \frac{1 \text{ cycle}}{7 \text{ bauds}} = 23.1428 \ \frac{\text{milliseconds}}{\text{baud}}$$

Two observations should be made here:

(a) Any dimension may be inverted (as was $\frac{7 \text{ bauds}}{1 \text{ cycle}}$) to suit the needs of the problem, and

DOCID: 3726612

(b) Inversion of the desired answer will result in a reciprocal value in the answer. For example, if the answer to the illustration were to be bauds per millisecond instead of milliseconds per bauds, the numerical value would also have to be inverted and solved to produce .04321 instead of 23.1428.

4. Any words or units not desired in the answer should be cancelled. However, several conditions should be noted:

(a) In order to cancel words they must appear both above and below the fraction bar.

(b) If matching words of the desired answer are cancelled the result will be altered so as to provide an undesired answer and

(c) All words that are not part of the desired answer must be cancelled leaving only the words matching those of the desired answer.

5. Some situations may require the inclusion of a conversion factor in order to express a given dimension in the proper units. Some of these are:

60 seconds per minute
60 minutes per hour
1 word per 6 cycles

Conversion factors involving milli, micro, kilo, and mega units per unit are not shown here because this only involves the basic conversion factor used with the appropriate power of ten. For example, we know that there are $\frac{60 \text{ seconds}}{1 \text{ minute}}$ by powers of 10 the student can derive that there are $\frac{60 \times 10^3 \text{ milliseconds}}{1 \text{ minute}}$

D. Dimensional Analysis and Powers of Ten

In the exercises that follow the student should substitute the powers of ten for the metric prefixes whenever possible. "$x10^3$" may be considered as a word just as if "kilo" appeared in its stead. If it occurs both above and below the fraction bar it may be cancelled just as any other unit in dimensional analysis.

DOCID: 3726612

With powers of ten many dimensional analysis problems can be solved without the use of paper. If division by a power of ten is to be performed, it is only necessary to move the power of ten from below to above the fraction bar and change the sign of its exponent.

To illustrate:

$$\frac{2 \text{ cycles}}{1 \times 10^{-3} \text{ seconds}} = \frac{2 \times 10^{3} \text{ cycles}}{1 \text{ second}}$$

EXERCISE V

1. Given: Oscilloscope sweep rate of 10 ms per centimeter

   20 cycles per 2 cm

   15 ms per baud

   Desired Answer: Frequency (cycles per second)

2. Given: 150 ms per cycle

   Desired Answer: Cycles per minute

3. Given: 7 bauds per cycle

   10 ms per baud

   Desired Answer: Words per minute

DOCID: 3726612

4. Given: Mark frequency -- oscilloscope sweep rate of 20 ms per cm;

10 cycles per cm

Space frequency -- oscilloscope sweep rate of 10 ms per cm;

2.5 cycles per cm.

Desired Answer: Frequency excursion.

5. Given: 132 ms per cycle

4 channels

Desired Answer: How many seconds are allotted to each channel?

6. Given: 7.5 bauds per cycle

22 ms per baud

Desired Answer: Words per minute.

7. Given: 22 ms per baud

368 operations per minute

Desired Answer: Bauds per cycle.

8. Given: 1 word equals 10 decimal digits

1 decimal digit equals 4 binary digits

100 wpm

Desired Answer: Milliseconds per binary digit.

9. Given: Oscilloscope sweep rate of 100 ms per cm.

50 cycles per cm

Desired Answer: Frequency.

10. Given: Oscilloscope sweep rate of 10 ms per cm

3 bauds per 6.5 cm

Desired Answer: Milliseconds per baud.

11. Given: 7.5 bauds per cycle

22 ms per baud

Desired Answer: Seconds per word.

12. **Given: 22 ms per baud**

    8 bauds per cycle

    Desired Answer: How many words are sent in 10 minutes?

13. **How far apart are the 1 second ticks of WWV when the Sonograph speed is 200 ms per inch?**

14. **Given: 5.5 ms per baud**

    132 ms per cycle

    4 channels

    Desired Answer: Bauds per channel.

15. **Given: 2.5 cm per baud**

    Oscilloscope sweep rate of 20 ms per cm

    24 bauds per cycle

    Desired Answer: Cycle length in milliseconds.

DOCID: 3726612

16. Given: 2.5 cm per baud

Oscilloscope sweep rate of 5 ms per cm

6 bauds per channel

Desired Answer: Words per minute.

17. An oscilloscope photo shows 12 sinewave cycles in 6 cm for a sweep rate of 2 ms per cm. Find the sinewave frequency.

18. An oscilloscope photo shows 1 pulse per 10 cm. Oscilloscope sweep rate is 5 microseconds per cm. Find pulse width.

19. Given: 5 ms per baud

140 ms per cycle

4 channels

Find the number of bauds per channel.

20. Given: 2.4 samples per cycle

8000 samples per second

Find: Frequency.

21. Given: 9.8 newtons per kilogram

kilogram per 2.2 pounds

pound per 16 ounces

Find: Ounces per newton.

22. Given: kilogram per 2.2 pounds

9.8 newtons per kilogram

Find: pound per newton.

DOCID: 3726612

CONCLUSION

Two points of emphasis have been considered separately in the preceding paragraphs: a method for obtaining correct numerical results and a method for obtaining correct dimensions. In any practical problem requiring a practical solution both the correct numeric and the correct dimension are necessary if the result is to be useful and worthy of the labor involved. The student is therefore encouraged to be painstakingly accurate.

MATHEMATICS FOR MODULATION has left many things unsaid due to the limitation of time and space in a text such as this and is to be considered only as a survey of the most used or most needed material. Therefore, if the student desires mathematical information beyond the scope of this survey, it is suggested that standard references, such as "Basic Mathematics for Electronics" by Nelson M. Cooke or "Basic Mathematics for Science and Engineering" by Andres, Miser, and Reingold, be consulted.

DOCID: 3726612

VI. PRACTICAL EXERCISES

1. ADD AND REDUCE:

$$\begin{array}{r} \frac{31}{52} \\ -\ \frac{7}{52} \\ \hline \frac{35}{52} \end{array}$$

2. SUBTRACT AND REDUCE:

$$\begin{array}{r} -\ \frac{17}{32} \\ -\ \frac{15}{32} \\ \hline \end{array}$$

MULTIPLY OR DIVIDE AS INDICATED, CANCEL WHENEVER POSSIBLE, AND LEAVE ANSWER IN LOWEST TERMS WITH PROPER SIGN:

3. $\frac{7}{15} \times (-1\frac{1}{2}) \times \frac{3}{7} =$

4. $2\frac{5}{8} \div 4\frac{2}{3} =$

5. $-15 \times \frac{2}{3} =$

6. $\dfrac{\frac{11}{12}}{\frac{5}{6}} =$

7. (a) ARRANGE IN COLUMNS AND ADD: 32.5, -.007, .046, -45, 3.25

   (b) EXPRESS THE ANSWER IN WORDS.

8. (a) ARRANGE IN COLUMNS AND SUBTRACT: -43.21 - (-78.2)

   (b) EXPRESS THE ANSWER IN WORDS.

DOCID: 3726612

CONVERT TO DECIMAL FRACTIONS:

9. $\frac{2}{5}$ =

10. $\frac{7}{32}$ =

CONVERT TO COMMON FRACTIONS IN LOWEST TERMS:

11. .425 =

12. .937 =

OBSERVING SIGN AND DECIMAL POINT, MULTIPLY OR DIVIDE AS INDICATED:

13. -4.62 x (-78.6) =

14. -901 x .008 =

15. -126 ÷ .072 =

16. -7.95 ÷ (-.636) =

17. Given: Oscilloscope shows 2.5 cm per baud

24 bauds per cycle

1.2 seconds per cycle

Find: Milliseconds per centimeter.

18. Given: 990 milliseconds per word

6 cycles per word

7.5 bauds per cycle

Find: Milliseconds per baud.

DOCID: 3726612

ANNEX B

OSCILLOSCOPE PRINCIPLES

Of the many instruments used in the analysis of various types of modulation, one of the most useful and versatile is the cathode ray oscilloscope. It is capable of measuring currents and voltages; displaying waveforms; measuring time such as wave periods and pulse widths; comparing and measuring frequency and phase; and is even able to measure high frequency voltages that could not be readily measured with a volt-meter. It should be observed, too, that spectrum analyzers, radar indicators, and television sets are all either adaptations or modifications of the cathode ray oscilloscope.

It is our aim to demonstrate to the student the mechanics of a cathode ray tube and to develop this into a simple but practical oscilloscope, showing how each of the block components in this simple arrangement relates to the oscilloscope and its desired operation. The concepts will then be broadened to show how a professional oscilloscope is simply a refined application of these fundamentals so that precise measurements may be made. This approach is used because it is believed that if the student knows what a particular control is for, he will never encounter any real difficulty in how to use it. Of the oscilloscopes available to operators, the TEKTRONIX 545 has proved to be the most versatile; detailed operating instructions, therefore, are given for the 545.

I. CATHODE RAY TUBES

The heart of any oscilloscope, irrespective of adaptation or modification, is the cathode ray tube (abbreviated CRT) (Figure 81).

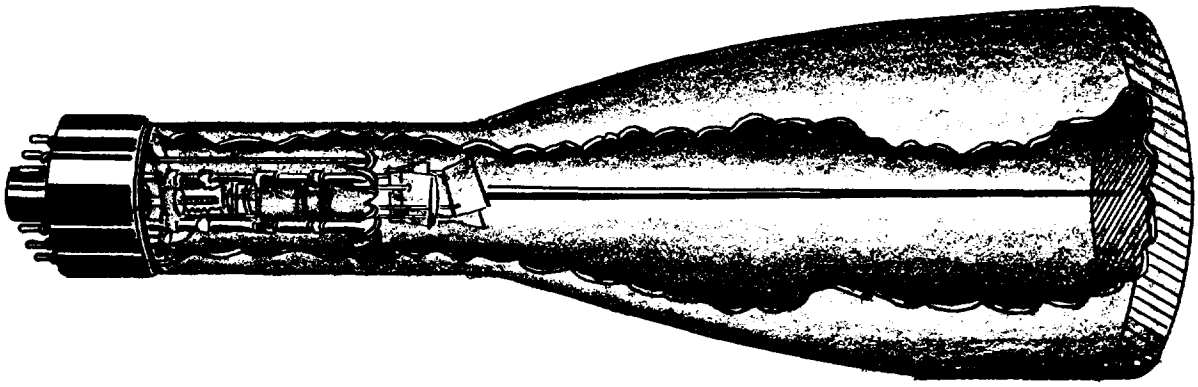


FIGURE 81. MODERN CATHODE-RAY TUBE OF ELECTROSTATIC-DEFLECTION TYPE.

DOCID: 3726612

The CRT is a hard vacuum tube designed to form and deflect an electron beam so as to describe on the face of the tube a pattern that is a function of the time variations of the input signals applied to the tube. The desired effect is accomplished in the three sections of the CRT namely, the electron gun, the deflection system, and the screen (Figure 82). (There are variations of construction for special purposes, such as the CRT's used to produce data encoding and pulse modulation rather than a visible waveform display, but these have no place in the present discussion.)

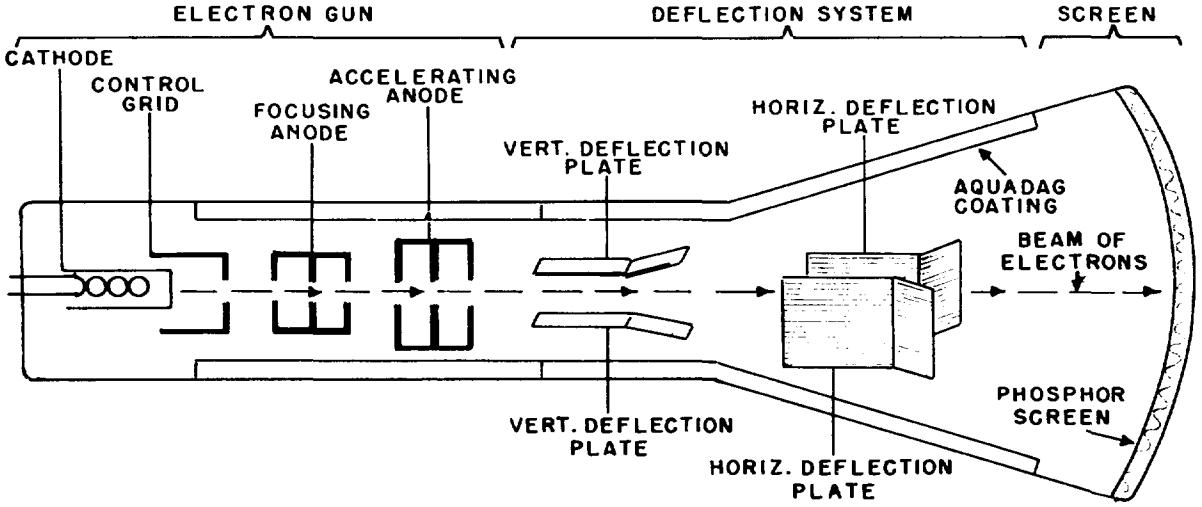


FIGURE 82. CONSTRUCTION OF CATHODE RAY TUBE (ELECTROSTATIC-DEFLECTION).

A. The Electron Gun

The first of these three sections describes the electron gun, the function of which is to provide a sharply focused stream of electorns. The basic elements of the electron gun are the cathode, the control grid, the focusing anode (first anode) and the accelerating anode (second anode) (Figure 83).

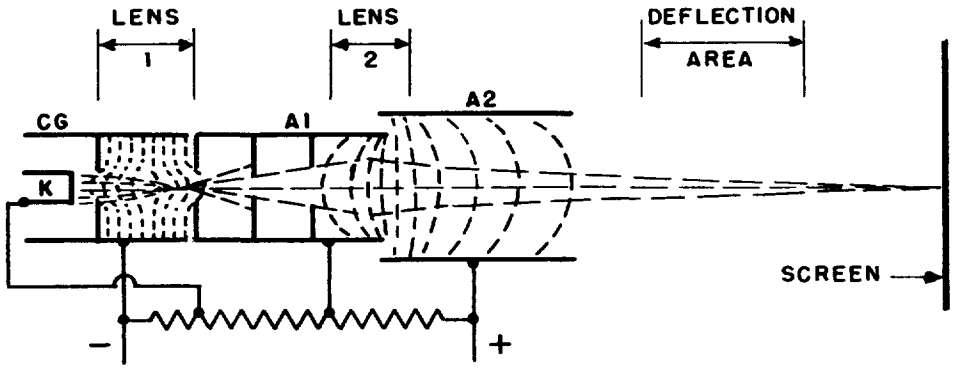


FIGURE 83. PRACTICAL ELECTROSTATIC FOCUSING SYSTEM SHOWING THE TWO SETS OF EQUIPOTENTIAL LINES THAT FORM THE LENSES OF A TRIODE ELECTRON GUN.

DOCID: 3726612

1. Cathode

The cathode is the indirectly heated type similar to those in other vacuum tubes. Its function is to emit a steady supply of electrons to the electron gun by means of thermionic emission.

2. Control Grid

The control grid is a hollow cylinder considerably larger than and surrounding the cathode. The control grid is closed on one end except for a small aperture through which emitted electrons may pass (Fig. 84).

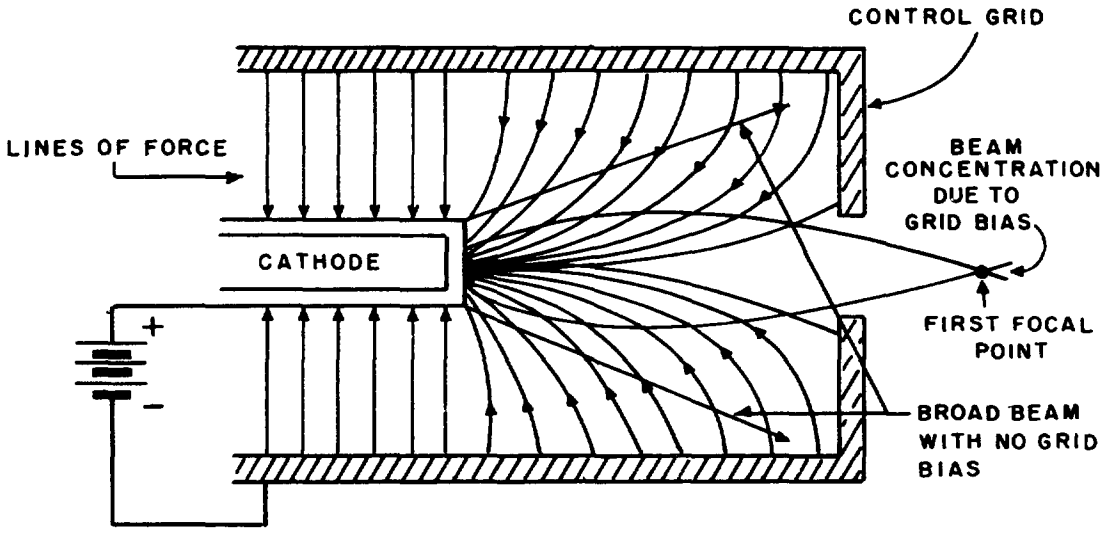


FIGURE 84. ACTION OF GRID IN CONCENTRATING ELECTRONS INTO A BEAM.

Since the grid is always negative with respect to the cathode it tends to repel emitted electrons back to the cathode or to concentrate them toward the aperture where the repulsion from the grid is considerably reduced. Making the control grid more or less negative will either reduce or intensify the number of electrons leaving the grid-cathode area by way of the aperture.

3. Focusing Anode

The next element in the sequence is the focusing anode, or first anode as it is sometimes called. The focusing anode has a cylindrical construction similar to that of the control grid with one exception. The anode is nearly closed on both ends, with only a small aperture in the ends through which the electron beam may pass (Figure 85).

DOCID: 3726612

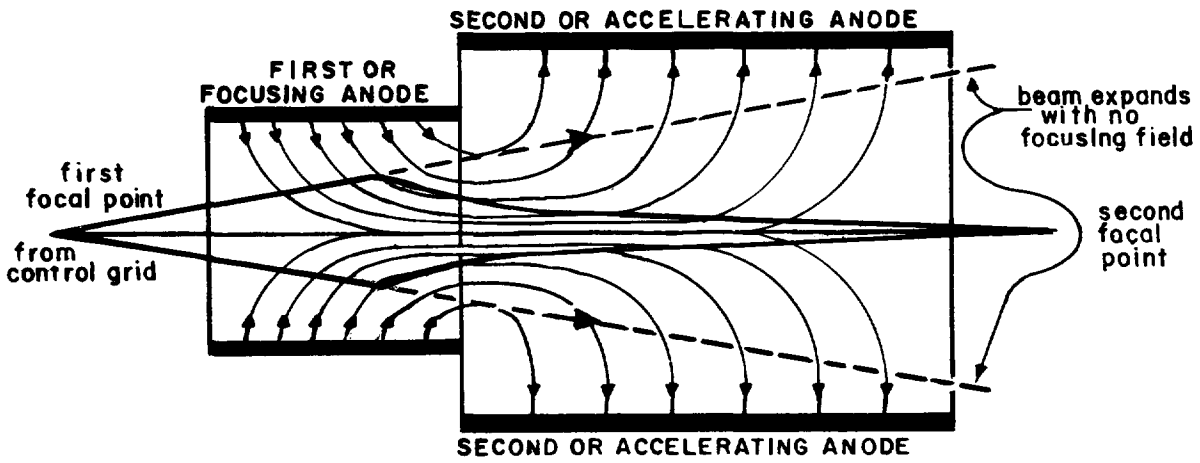


FIGURE 85. FOCUSING ACTION.

This element is many volts positive with respect to both the cathode and the control grid. The resulting force of attraction not only attracts the electrons passing through the control grid aperture, but is actually felt by electrons as they are emitted from the cathode surface. The combined effect of the repulsion from the negative grid and attraction from the highly positive first anode constricts the electrons passing through the grid aperture into a very narrow, highly-concentrated, pencil beam. It is this beam of electrons that constitutes the cathode ray. The combined force is so great that the electrons leaving the aperture are made to converge on a point, cross over one another and then diverge. This cross-over is very similar to the cross-over at the focal point when light passes through an optical lens. So similar are the two effects that the cross-over is called the first focal point. The focusing anode, together with the control grid, constitutes the first lens system. Similar to moving a lens in an optical system, varying the voltage on the focusing anode will move the focal point and hence will vary the focus of the beam. Any electrons that diverge too far from the axis of the beam will be unable to pass through the focusing anode aperture and will therefore be removed from the beam by the anode. This insures that the cathode ray will be kept very narrow.

4. Accelerating Anode

The fourth and final element of the electron gun is the accelerating anode, sometimes called the second anode (Figure 85). Applied to this element is a large positive voltage greater than that applied to any other

DOCID: 3726612

element in the electron gun. The shape of this second anode is very similar to that of the first anode, although the cylinder is somewhat shorter. The purpose of the accelerating anode is to attract the electrons, thus imparting to them a high velocity, and yet let them pass through the anode by way of its apertures. The acceleration is caused by the very high positive voltage. The small apertures serve the same purpose as before, namely, to remove straying or divergent electrons. The difference in voltage between the focusing anode and the accelerating anode causes the beam to converge at a second focal point where the electrons strike the CRT screen. The interaction of the first and second anodes produces this focusing action, thus forming the second lens system.

B. The Deflection System

After the electron gun has performed its functions of forming the emitted electrons into a narrow beam, accelerating them to a high velocity, and focusing them to a sharp point at the face of the CRT, it is still necessary to move the beam about to describe the desired pattern. The beam may be deflected electrostatically or magnetically, but only the electrostatic deflection system will be discussed here.

After the beam leaves the electron gun, it passes between two pairs of metal plates (Figure 86).

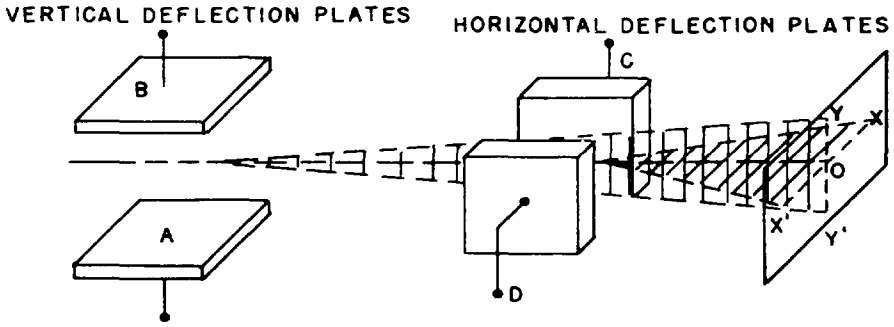


FIGURE 86. DEFLECTION PLATES AND SCREEN OF A CATHODE RAY TUBE.

One pair is placed in the vertical plane (X or horizontal deflection plates) and the other is in the horizontal plane (Y or vertical deflection plates). The effect utilized in the electron gun--that an electron is repelled by a negative charge and attracted by a positive one--is applied again in the deflection system. By making plate A negative or positive with respect to plate B, the electron beam can be deflected up and down. (Observe points Y, O, and Y′.) By making plate D negative or positive with respect to plate C, the beam can be made to move back and forth. (Observe points X, O, and X′.) If voltages are applied to both sets of plates, the beam can be deflected to any position.

DOCID: 3726612

C. The CRT Screen

The glass envelope of the cathode ray tube is enlarged at the end opposite the electron gun to form a circular screen. The inside of the screen is coated with a material called a phosphor which gives off light when struck by electrons. When the electron beam is formed to a point just as it strikes the screen a pinpoint of light will be produced. The phosphor is both fluorescent and phosphorescent, i.e., it will give off light both when it is excited by the electron stream, and after the excitation is removed.

It is this latter quality that gives the screen persistence permitting the path of the beam to be seen for a short time after the beam has moved on. Were it not for the persistence of the CRT screen, the oscilloscope trace would appear as a moving dot instead of a continuous line.

Associated with the screen, but not part of it, is a coating of carbon particles on the side walls of the tube known as the aquadag coating. This coating is connected to the high voltage source that supplies the accelerating anode and therefore is at a high positive voltage. The coating is intended to collect electrons knocked out of the screen by the impact of the high energy electron beam.

D. Cathode Ray Tube Operation

The overall operation of the cathode ray tube may be summarized by associating its elements with the controls and signals applied to them.

1. Intensity vs Control Grid

The intensity of light emitted from the screen phosphor is directly dependent upon the energy of the electron beam, i.e., the greater the beam energy, the greater the intensity. The energy of the beam on the other hand is a function of two things: the quantity of electrons and the velocity of the electrons. It naturally follows that if either the quantity or velocity of the electrons could be controlled at will, the intensity of the screen display could also be controlled. For reasons to be explained it is unwise to change the velocity. The control grid, however, not only concentrates the electrons emitted from the cathode into a narrow beam but also determines the quantity of electrons composing that beam. The control grid, therefore, is the element best suited to controlling intensity. An intensity control on an oscilloscope, then, is simply a control that regulates the number of electrons that the control grid will pass.

DOCID: 3726612

2. Focus vs Focusing Anode

Both the focusing anode and the accelerating anode have a definite effect on the focus of the electron beam. It has been stated that the accelerating anode controls the velocity of the electrons and thus would be unsuited as the controlling element in regards to focus. The focusing anode, however, is best suited to focusing the beam because the voltage difference between the focusing anode and the control grid (first lens system) determines the first focal point; the voltage difference between the focusing anode and the accelerating anode (second lens system) determines the second focal point. The focusing anode exercises control on the focus at two points since it is part of two lens systems. One can readily see, then, that the focus control on an oscilloscope is simply a means of adjusting the voltage on the focusing anode.

3. Positioning vs Deflection Plates

Two pairs of deflection plates are located between the electron gun and the screen (see Figures 81, 82, and 86). The electron beam must pass between each pair of these plates on the way to the screen. As the beam passes through the deflection plates it will experience a force that will bend or deflect it if any voltage is present on the plates.

(a) Deflection Sensitivity. The amount of deflection observed on the screen will be proportional to the bend or deflection angle. The deflection angle, on the other hand, is directly proportional to the voltage applied to the plates. The ease of deflection is described in terms of deflection sensitivity which is expressed as deflection/volt (e.g., 2 millimeters/volt or 1 inch/volt, etc.). The reciprocal of deflection sensitivity is frequently used (as volts/cm) and is called the deflection factor. Either pair of deflection plates could be used for vertical or horizontal deflection. Since the plates nearest the electron gun have the greatest deflection sensitivity, these are usually used for vertical deflection.

(b) Positioning with Fixed Voltages. Suppose that both the vertical and horizontal deflection systems of a certain CRT have the same deflection sensitivity, namely, .01 inch per volt. This will allow 10 volts to produce 1/10 inch deflection, or 20 volts to produce 1/5 inch deflection. Figure 7 shows the progressive position of the beam in the horizontal plane when voltage is applied to the horizontal deflection plates in 20 volt steps from 0 to 100 volts and back through 0 to -100 volts. Whenever the voltage difference between a pair of plates is zero, there is no deflection. (Had the battery been applied

DOCID: 3726612

to the vertical deflection plates the beam would have stopped at progressive positions along the vertical axis.)

If voltages were applied to both pairs of plates at the same time the beam could be brought to rest at any desired point. Observe Figure 87A and Figure 87B. Figure 87A shows the position of the beam in the horizontal plane with +60 volts applied to the horizontal deflection plates. Figure 87B shows the beam displaced along the horizontal axis as before, and simultaneously displaced along the vertical axis by a voltage applied to the vertical plates.

(c) Positioning with Varying Voltages. If a varying or alternating voltage were applied to the vertical deflection plates as shown in Figure 88, the beam would trace a continuous path through the points labeled A, B, C,...G, H, I.

While leaving the sinusoidal voltage on the vertical deflection plates, suppose we apply a voltage to the horizontal deflection plates that varies linearly with time. Also, suppose that the time required to sweep the beam across the screen horizontally is equal to the period of one cycle of the voltage on the vertical deflection plates. The result would be that one cycle of the vertical signal will be traced on the CRT screen as illustrated in Figure 89.

DOCID: 3726612

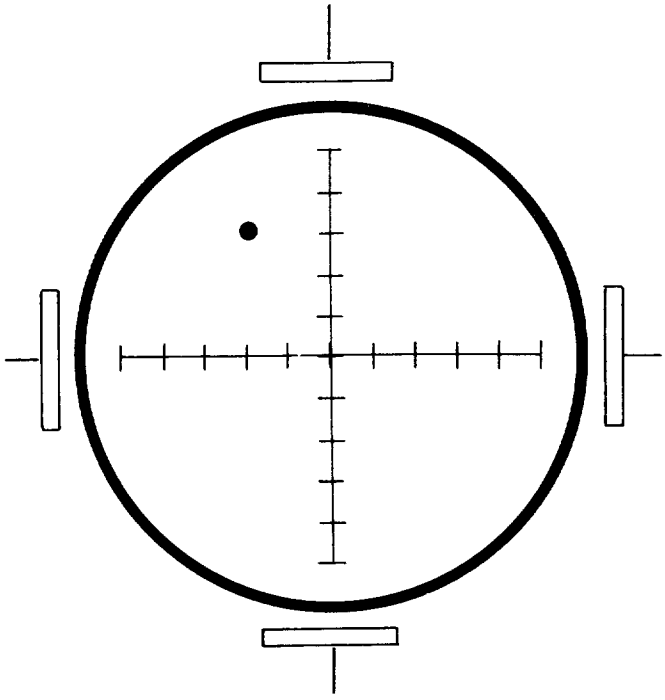


FIGURE 87B. DEFLECTION WI TH +60 VOLTS ON HORIZONTAL DEFLECTION PLATES AND +40 VOLTS ON VERTICAL DEFLECTION PLATES (VOLTAGE SOURCES NOT SHOWN).

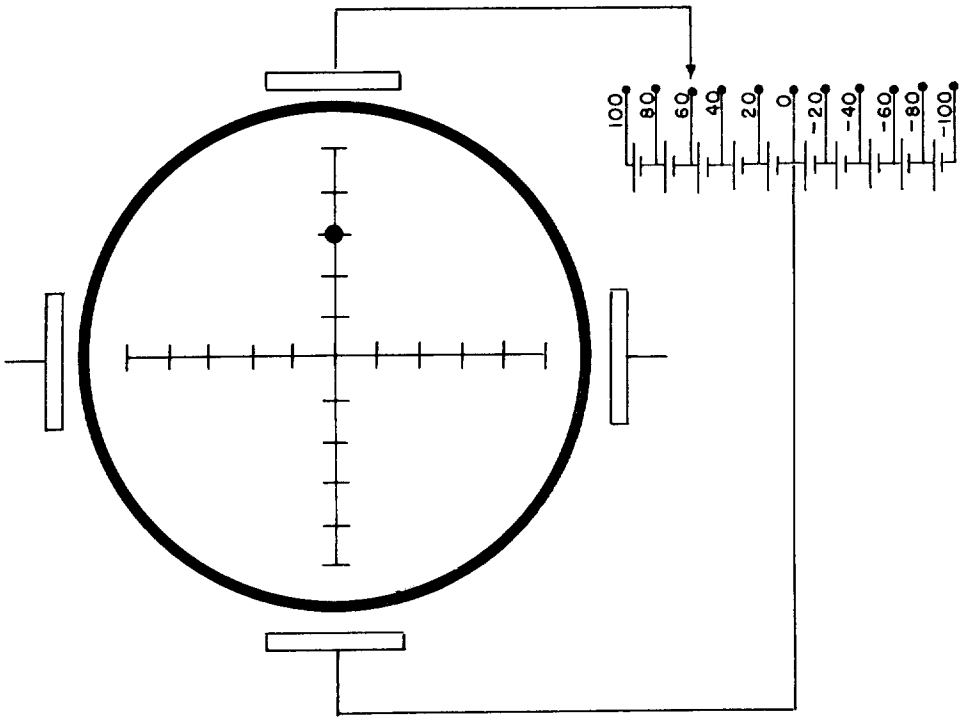


FIGURE 87A. DEFLECTION WITH +60 VOLTS ON HORIZONTAL DEFLECTION PLATES.

DOCID: 3726612

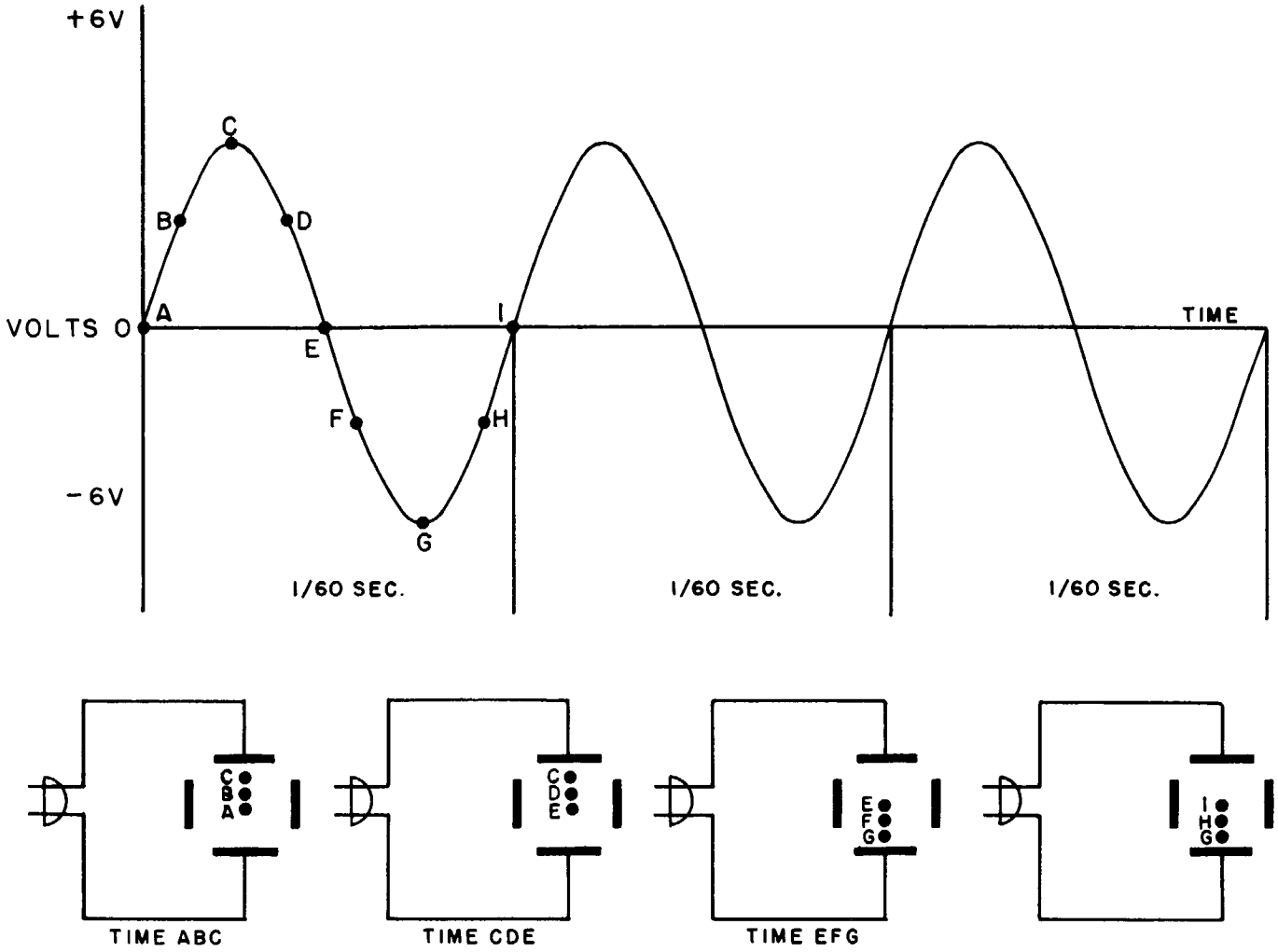


FIGURE 88. DEVELOPMENT OF A WAVEFORM ON THE SCREEN.

DOCID: 3726612

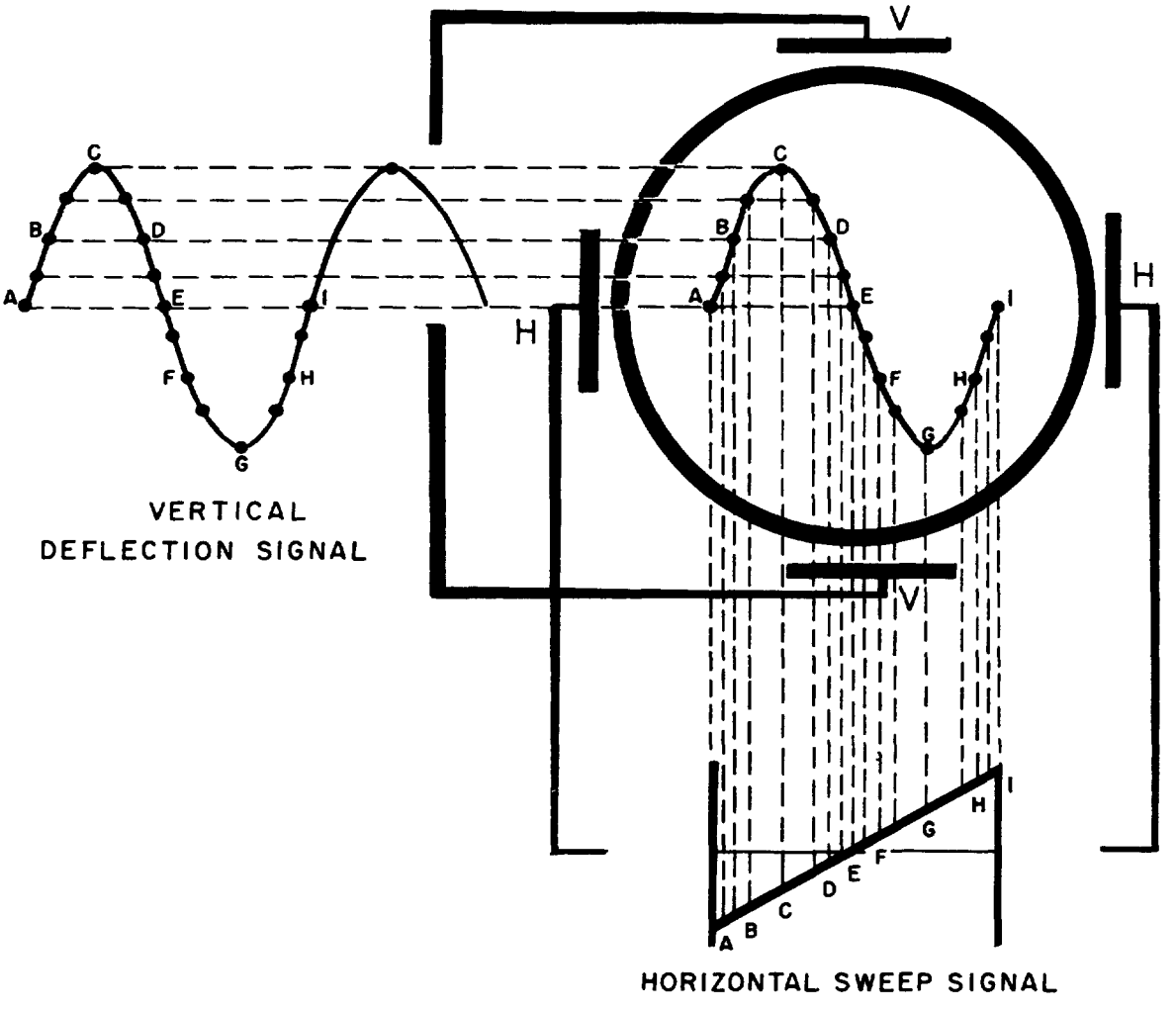


FIGURE 89. LINEAR TIME-BASE USED TO DEVELOP A WAVEFORM.

DOCID: 3726612

4. Screen Luminescence vs Screen Material

The color of the trace on the face of a cathode ray tube should not lead the student to believe that electrons are green or yellow or any other color for that matter. The color of the glow is determined by the screen material. For oscilloscopes, a phosphor giving a bright green color is usually employed in the CRT since this color is easily obtained and has good visual quality. For photography, a deep blue glow is more suitable because of its greater ability to produce photochemical reactions. By using a suitable mixture of phosphor, a glow approaching white may be obtained. This white glow is preferred for television screens because it causes the pictures to present a more pleasing appearance.

Various phosphorescent materials exhibit different rates of decay of brightness after the excitation has been removed. We have already defined this phosphorescent after-glow as persistence. Tubes are available with either long or short persistence. Those with long persistence usually glow in two colors--a short blue fluorescent flash during excitation followed by a long persistence of phosphorescent yellow or orange. Such screens are useful for study of extremely brief phenomena, for example, radar information. For general oscillographic purposes a screen with medium persistence is used.

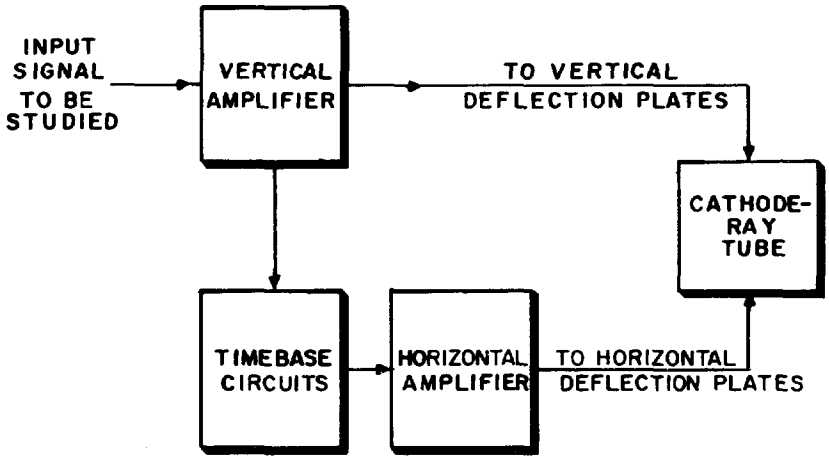


FIGURE 90. SIMPLIFIED BLOCK DIAGRAM OF THE BASIC OSCILLOSCOPE.

DOCID: 3726612

II. THE BASIC OSCILLOSCOPE

The oscilloscope is an electronic device which provides a means for plotting a visual graph on the screen of a cathode ray tube. The CRT is, in a sense, a simple oscilloscope; however, even the most basic oscilloscopes for any practical usage are somewhat more elaborate. The refinements consist primarily of amplifiers for increasing the sensitivity of the cathode ray tube to small signals, and a sweep generator for supplying a stable and controllable time reference signal to the horizontal deflection plates.

Figure 90 presents the block diagram of the signal circuits of a basic oscilloscope. The power source, although essential to the operation of every part of the oscilloscope, is not a signal circuit and therefore is not included in this discussion. The components of interest are the vertical deflection amplifier, the time-base generator, the horizontal deflection amplifier and the cathode ray tube.

A. Vertical Deflection Amplifier

Generally the deflection sensitivity of a CRT is not great enough to give an adequate display of small amplitude signals, so in most instances it is first necessary to increase the amplitude of the input waveform. Inserting a vertical deflection amplifier between the input terminals of the oscilloscope and the vertical deflection plates has the effect (so far as the operator is concerned) of increasing the deflection sensitivity of the cathode ray tube. Gain controls on the vertical deflection amplifier then make it possible to vary the effective deflection sensitivity so that the input waveform will deflect the beam (and hence the trace) the desirable amount.

B. Time-Base Generator

If the input signal is observed on the face of the cathode ray tube without any signal applied to the horizontal deflection plates, all that will be seen is a spot moving up and down, tracing a vertical line like that shown in Figure 88. To follow the waveform of the vertical deflection signal with respect to time, it is necessary to have a time-base against which the signal can be plotted. Such a time-base supplied to the horizontal deflection plates fulfills the same function as the time-base in Figure 1 of Waveform Analysis and is the same as the electronically derived time-base described previously in this lesson with respect to Figure 89.

The time-base generator produces a voltage that increases linearly with time (Figure 89). After amplification this voltage is applied to the horizontal deflection plates of the CRT. This voltage causes the electron beam to sweep horizontally across the face of the CRT at a uniform rate and then return rapidly to its initial position. This action, of course, is repeated continuously. (In passing,

DOCID: 3726612

it should be called to the student's attention that the repetition of the linearly increasing sweep voltage and its rapid return to zero constitutes a sawtooth wave train.) Since the time-base voltage is a linear function of time, the input signal is shown on the oscilloscope as a graph of voltage plotted against time.

Various sweep rates are available as selected by coarse sweep rate and fine sweep rate controls. An additional control used with this unit is a sweep synchronization or sync control.

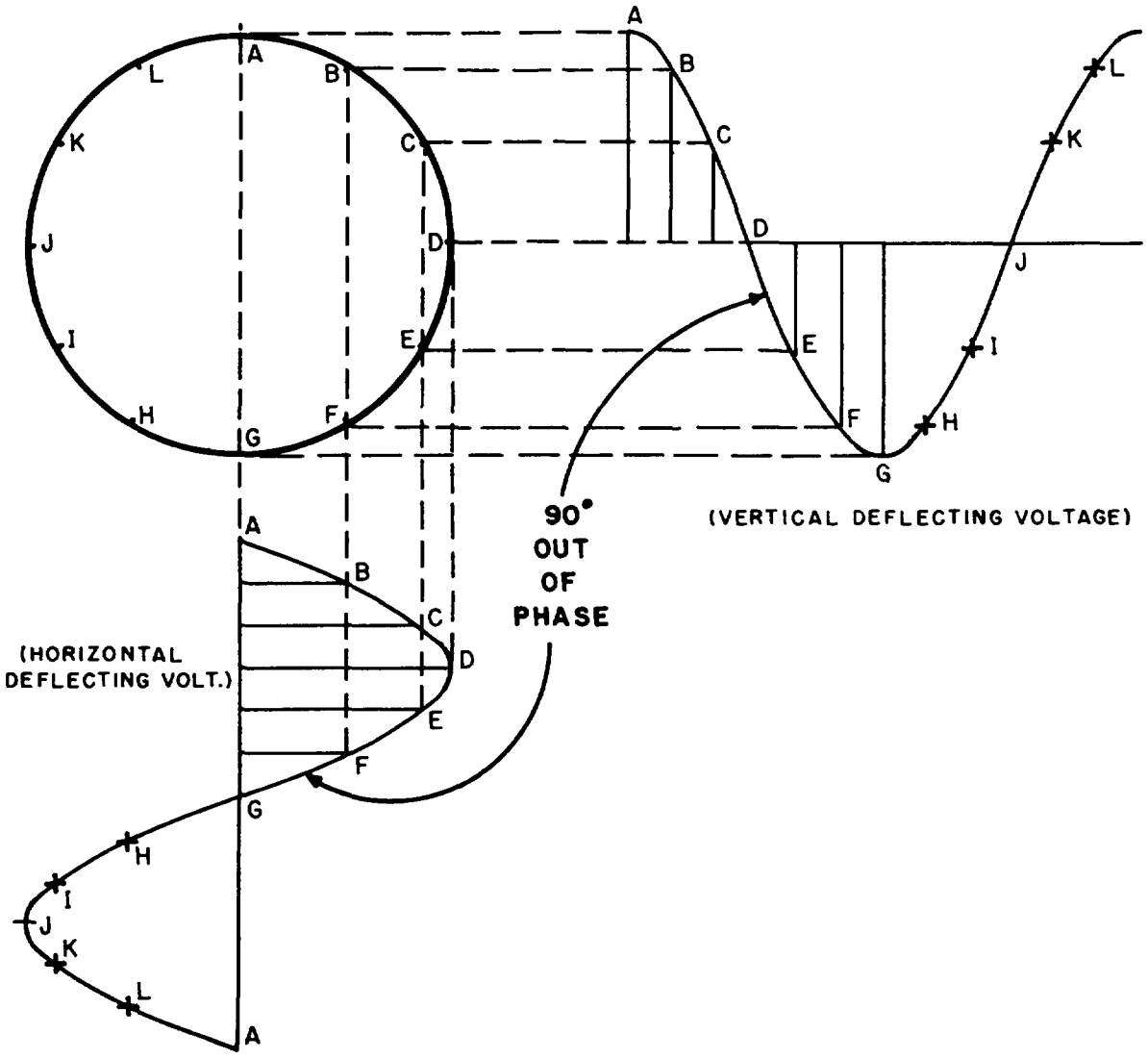


FIGURE 91. LISSAJOUS PATTERN RESULTING FROM TWO SINEWAVES 90° OUT OF PHASE.

DOCID: 3726612

C. Horizontal Deflection Amplifier

As waveforms fed to the vertical deflection plates often need to be amplified, so signals to be applied to the horizontal deflection plates may need amplification to improve the overall deflection sensitivity of the horizontal system. It is the purpose of the horizontal deflection amplifier to perform the latter operation.

Generally, the signal that is amplified by this unit is the sweep voltage from the time-base generator. However, there are occasions when it is desirable to put some other waveshape, or a wave from some other source, on the horizontal deflection plates. For example, when it is necessary to make frequency or phase measurements by means of Lissajous patterns, a sinewave sweep is used for a time reference. The Lissajous pattern resulting from two sinewaves of like frequency but differing by 90 degrees in phase is shown in Figure 91. Choice of internal or external sweep is made by means of a sweep selector switch that determines the input to the horizontal deflection amplifier. The only other control associated with this unit is the horizontal gain control.

D. The Complete Oscilloscope

It should be apparent by now that an oscilloscope is only a cathode ray tube to which certain accessories and refinements have been added. A moment of reflection will make it evident that each of these refinements is both a necessary and logical addition in developing a useful and versatile instrument. In the complete oscilloscope, it is common practice to group the controls according to their functions so that their placement on the front of the instrument is compatible with the scheme of operation (Figure 92).

This discussion would not be complete without a brief description of how the controls are applied to govern oscilloscope operation in typical general purpose viewing. The method is not rigid because a particular model of instrument or a particular application may dictate certain exceptions. Some of these exceptions will be realized in this lesson when the student makes a detailed study of a typical laboratory oscilloscope. The immediate description is a guide only and will apply in a general way to most oscilloscopes.

1. Preliminary Settings

Before an oscilloscope is turned on, the intensity control should always be turned to minimum intensity (normally counterclockwise) because if an intense spot is allowed to remain stationary on the screen there is danger of burning the phosphor coating.

DOCID: 3726612

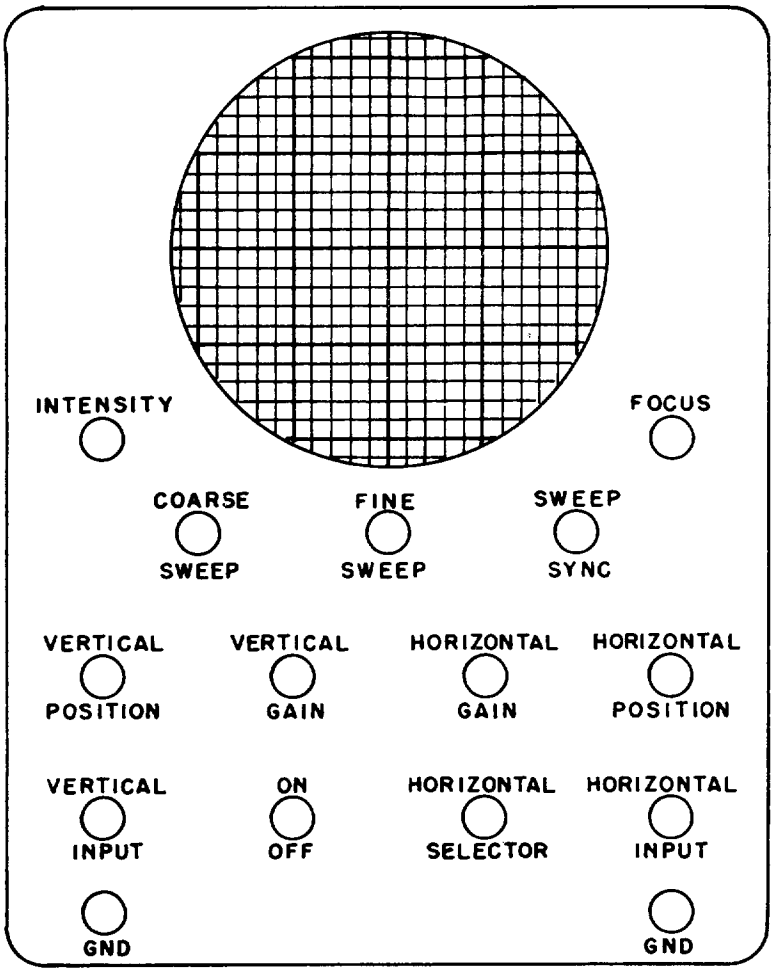


FIGURE 92. BASIC OSCILLOSCOPE

After the instrument has been turned on and sufficient time allowed for warm-up, it should be determined that the sweep generator (either internal or external) is properly set to produce a trace (though not yet visible) across the screen at some arbitrary sweep rate.

2. Adjusting the Trace

The intensity and focus controls are now adjusted in turn to obtain the clearest and sharpest trace possible. Although operation of the sweep circuits reduces the likelihood of burning the phosphor, caution must be exercised when adjusting the intensity as the high energy of too bright a trace can destroy the screen material. Adjustment of the intensity may require a slight readjustment of the focus control since any variation in intensity tends to defocus the beam.

DOCID: 3726612

3. Positioning the Trace

The vertical positioning control is next used to center the trace up and down. Following this, if the screen is provided with a scale, grid lines, or a graticule, the horizontal positioning and horizontal gain controls are adjusted in conjunction so that the trace originates at the left-hand limit of the scale and terminates at the right-hand limit. If there is no scale, the horizontal gain and positioning are adjusted so that the trace extends across the flat part of the screen (but not beyond it) and is centered left and right.

4. Vertical Gain Adjustments

These adjustments have prepared the oscilloscope for application of the input signal which must be connected between the vertical input terminal and the ground terminal. After applying the input signal, the vertical gain control is varied until the signal deflects the trace vertically over a little less than the middle two-thirds of the screen unless other convenient limits are described by a graticule or scale. In the absence of a graticule or scale, vertical deflection much in excess of the limits suggested here tends to result in defocusing and distortion of the waveform.

5. Regulating Sweep Rate

It will be remembered that to avoid the risk of burning the screen the sweep was arbitrarily set just to obtain a trace. Final adjustments relate to setting the sweep controls so a useful presentation of the input may be observed. It is a matter of choice as to the number of cycles of the vertical input that should be displayed across the screen. For good waveform analysis it is best to display no less than two cycles (three are preferred); but in most cases the display of five or six cycles is entirely too many. Special instances will dictate any special treatments.

When the period of the sweep is equal to the period of the vertical input, one cycle of the input waveform occurs in the same time as one sweep of the trace. Therefore, in order to display one cycle of the input waveform the sweep rate must be equal to the vertical input frequency. If, for example, three cycles of the input waveform are to be obtained as suggested above, the period of the sweep must be three times as long as the period of the input. In terms of frequency the sweep rate must be set to one-third the frequency of the input. A rule may be formulated from this which may be expressed as:

DOCID: 3726612

$$\text{sweep rate} = \frac{\text{wave frequency}}{\text{cycles displayed}}$$

Whatever the desired display, the coarse sweep rate control must be set to approximate the rate indicated above. The fine sweep rate control is then adjusted until the desired pattern becomes approximately stationary. The sweep synchronization control is then used to "lock" the sweep rate to the input frequency so that the presentation is held absolutely stationary.

III. A PROFESSIONAL LABORATORY OSCILLOSCOPE

A typical professional quality laboratory oscilloscope capable of displaying fast-rising waveforms is the TEKTRONIX 545 shown in Figure 93. Fast rise-time amplifiers and a high accelerating voltage applied to the cathode ray tube enable this oscilloscope to be used in the analysis of pulse trains. The measurements and analysis of such wave trains are further facilitated by accurately calibrated sweep time and vertical deflection factor (deflection factor as defined by the Institute of Radio Engineers, 50 IRE 7.S1). The availability of interchangeable plug-in preamplifiers for the vertical input increases the flexibility and adaptability of such an instrument to meet widely varied laboratory application. No attempt is made here to elaborate on or even mention all the controls on this model, but only to explain the purpose and use of the controls that students of this training program will need.

The manufacturer's literature divides the 545 system into five major divisions: vertical deflection, main sweep, delaying sweep, sweep amplifier, and horizontal display, but we will use a somewhat different break-down in order to show how this advanced system still conforms to the block components of a basic oscilloscope.

A. Vertical Deflection Amplifier System

The vertical deflection system of the 545 oscilloscope consists of an amplifier built into the main unit of the instrument itself and a preamplifier which plugs into the oscilloscope. This preamplifier is easily removeable and interchangeable with other preamplifier types. When mounted in the oscilloscope the plug-in unit operates as a part of the instrument's vertical deflection amplifier system with all the controls of this system located on the face of the plug-in preamplifier. For this reason our discussion of the vertical deflection amplifier will deal with the preamplifier only. The immediate discussion is based on the TYPE K preamplifier shown in Figure 94.

DOCID: 3726612

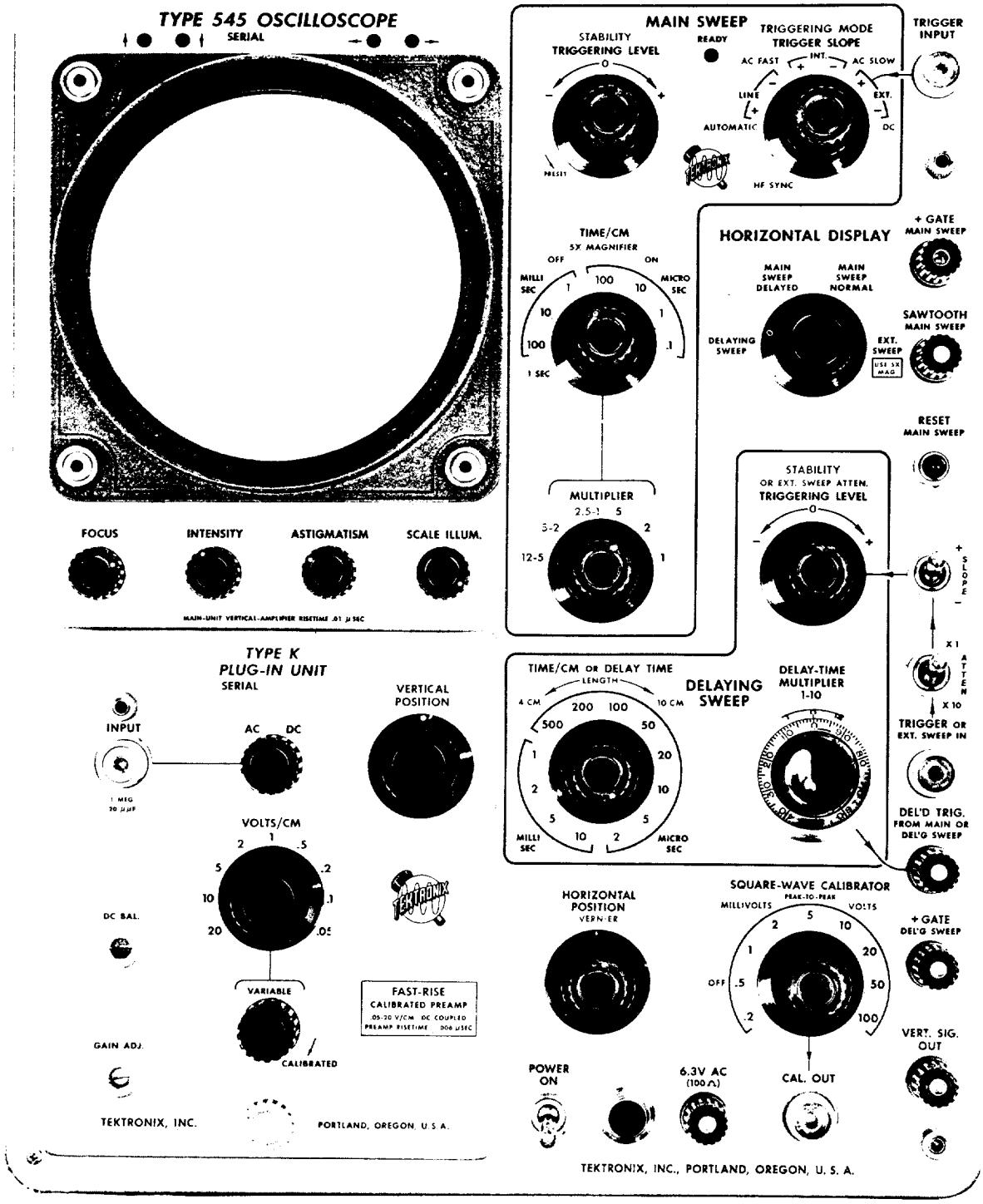


FIGURE 93. THE TEKTRONIX 545.

DOCID: 3726612

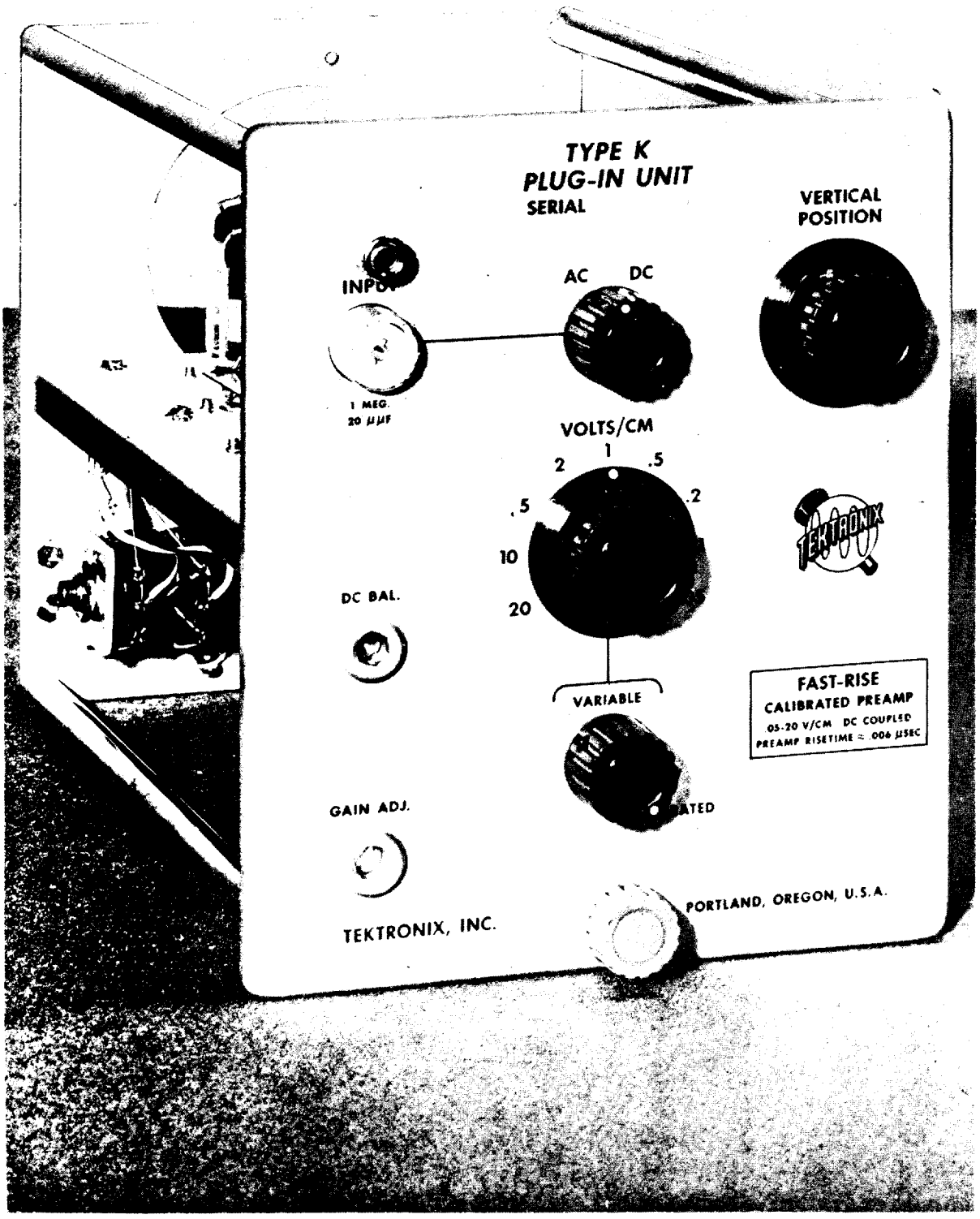


FIGURE 94. THE TYPE K PREAMPLIFIER.

DOCID: 3726612

1. Input

By means of a probe and a UHF coaxial cable provided with the equipment, the input signal (which may be either ac or dc) is applied to the vertical deflection amplifier through the coaxial connector in the upper left-hand corner of the TYPE K plug-in unit. The input impedance as seen through the P410 probe is 10 megohms shunted by 7.5 picofarads. The input voltage to this unit should not exceed 600 volts with the dc voltage and the ac peak voltage combined. The pass band that this amplifier provides to the input signal is dc to 30 megacycles per second.

2. Controls

All the operator controls relating to the vertical deflection amplifier systems are mounted on the face of the preamplifier. These controls are the AC-DC switch, vertical position control, volts per centimeter switch, and variable gain control.

(a) The AC-DC Switch. This control is provided to change the type of input coupling to the preamplifier. In the DC position both the ac and dc components of a wave are allowed to pass through the vertical deflection amplifier system, but in the AC position the dc component is blocked.

(b) The Vertical Position Control. The name of this control makes its function fairly obvious. Its purpose is to vertically center the oscilloscope trace on the screen.

(c) The Volts per Centimeter Switch. It is often necessary to adjust or vary the vertical deflection factor of the oscilloscope to accommodate signals of various strengths. This capability is accomplished by means of the Volts/CM switch which is an extremely accurate step attenuator. When the variable gain control (to be discussed) is in the "calibrated" position the calibration of each of the discrete steps of the volts/cm switch is an accurate reading of the deflection factor of the oscilloscope. The basic deflection factor of the 545 is .05 volts/cm but the volts/cm switch permits this to be reduced to any of eight other levels.

(d) The Variable Gain Control. This control adjusts the gain of the preamplifier and therefore adjusts the gain of the whole vertical deflection system. It varies the gain over a ratio of about 2 and 1. When turned to the extreme clockwise position, the variable gain

control calibrates the volts/cm switch to read out the deflection factor directly.

(e) Other Adjustments. The screwdriver adjustments appearing on the face of the preamplifier are not for the use of operator personnel enrolled in this course but are for use of maintenance personnel only.

B. Time-Base System

The time-base system in the 545 is considerably more complicated than those of the basic oscilloscope but the ultimate purpose and function are identical. The complexity is in the operator's favor, however, since it extends the versatility of the oscilloscope.

Just as it was possible to display a sweep from either an internal or external source with the basic oscilloscope, this same possibility exists with the 545; however, the latter permits the use of three different internal sweeps.

1. Horizontal Display Selector

Horizontal deflection of the electron beam by the three internal sweeps or an external sweep is arranged for by selecting one of the four positions of the Horizontal Display Selector located to the extreme right of the CRT screen (see Figure 95). These four positions are Main Sweep Normal, Delaying Sweep, Main Sweep Delayed, and External Sweep.

(a) Main Sweep Normal. This operation is most useful for general purpose observation and wave period measurements. Main Sweep Normal may be either triggered or synchronized and may be operated in any of five modes (para. 2 (b)).

(b) Delaying Sweep. Although delaying sweep will not ordinarily be used by students of this course it may be used as a quick check on the main sweep operation. It may be triggered externally by an external signal applied to the Trigger Input connector or the trigger may be taken from the observed waveform by connecting the Vertical Signal Out connector to the Trigger Input connector.

(c) Main Sweep Delayed. When main sweep delayed is used both the Main Sweep circuits and the Delaying Sweep circuits show in the operation. Among the several capabilities that this operation offers are: (1) to select a particular portion of the wave train observed

DOCID: 3726612

on Delaying Sweep and expand this portion on Main Sweep to fill the entire horizontal display by switching over to Main Sweep Delayed, (2) to permit jitter-free displays even when the observed waveform jitters by keeping the Main Sweep inoperative for some predetermined time delay and then permitting it to be triggered for a single display and (3) to manually obtain single sweeps for purposes of making photographs of waveforms.

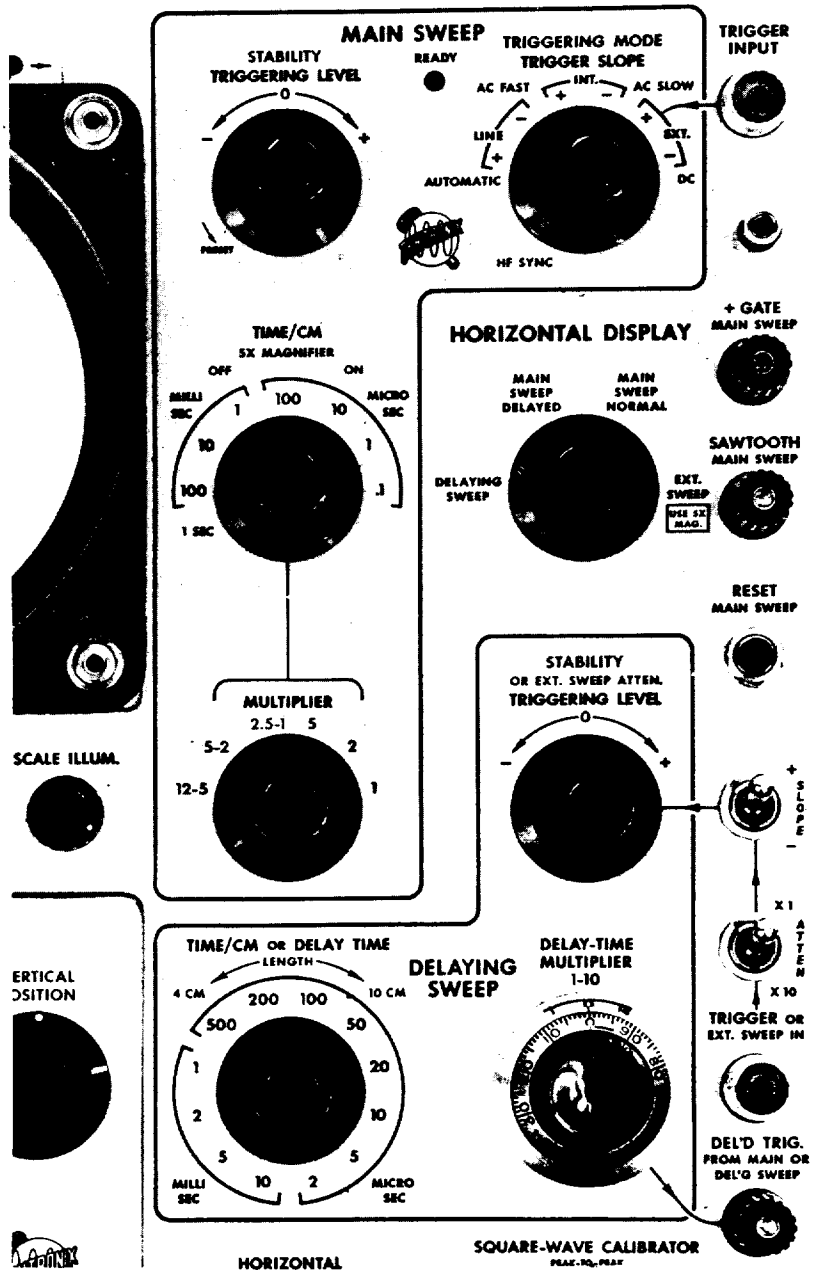


FIGURE 95. TIME-BASE SYSTEM CONTROLS.

DOCID: 3726612

(d) External Sweep. Just as it was possible to provide horizontal deflection in the basic oscilloscope from some source other than the time base generator, this same capability exists with the 545 by supplying the desired signal to the "Trigger or External Sweep In" connector and switching the Horizontal Display Switch to the External Sweep position. External sweep is extremely useful for making frequency measurements from Lissajous patterns.

2. Main Sweep.

The controls in the main sweep section of the 545 oscilloscope shown in Figure 96 may be considered in two categories, namely, time controls and triggering controls. The time controls are the time per centimeter control, the multiplier, and the 5X Magnifier. The triggering controls are Triggering Mode, Trigger Slope, Stability, and Triggering Level.

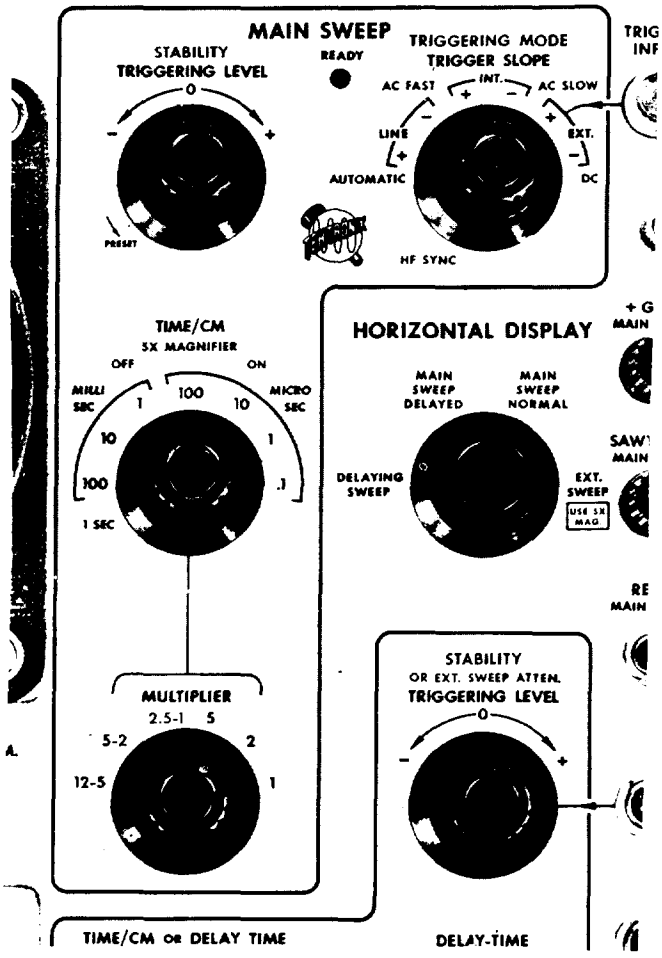


FIGURE 96. MAIN SWEEP SECTION CONTROLS.

(a) Time Controls. The horizontal trace produced as the electron beam is swept across the CRT screen normally traverses a length of 10 centimeters. The beam may sweep across the screen so slowly as to take 50 seconds and will appear as a scarcely moving dot. On the other hand it may make the entire sweep of 10 centimeters in as little as one five millionth of a second (200 nanoseconds) in which case the dot is moving so fast only its path can be seen as a luminous trace across the screen. The rate of speed of the electron beam may be measured in exactly the same manner as we describe the speed of an automobile in miles per hour, that is, distance per unit of time. For our measurements we are more interested in the time required for the unit distance than we are in distance per unit of time (note that one is the reciprocal of the other), therefore the appropriate control is calibrated in terms of sweep time rather than in terms of sweep rate.

(1) Time per Centimeter Control. The rate at which the beam is swept across the screen is determined by the setting of the black control labeled Time/Cm. Starting at the extreme clockwise position and stepping counterclockwise, each step is seen to represent 10 times as much time as the previous position. To avoid cumbersome readings the calibration is broken down into microseconds, milliseconds, and seconds.

(2) Multiplier. Beneath the Time/Cm Control is another black knob labeled Multiplier. It has six positions: three marked in black which indicate factors by which the Time/Cm settings are to be multiplied to obtain the main-sweep rate, and three positions marked in red which provide a continuously variable (uncalibrated) main sweep rate control by means of an associated red knob.

(3) 5X Magnifier. This control is the red knob concentric with the Time/Cm control. When the 5X Magnifier is in the ON position the center one-fifth of the trace is expanded to occupy the entire trace. Actually this increases the sweep rate such that the .1 microsecond setting of the Time/Cm control represents .02 microseconds per centimeter (20 nanoseconds/Cm). Normally the 5X Magnifier control will be left in the OFF position.

(b) Triggering Controls. It is normally desirable to obtain a stable display of an observed waveform. A stable display can only occur when the start of each sweep of the electron beam is synchronized with the occurrence of some specific point on the observed waveform. On the 545, synchronization is acquired by initiating each sweep by a triggering or synchronizing pulse which may be derived from the input signal or from some external source. Synchronizing the sweep by means of a synchronizing pulse in this manner is called triggering. The triggering controls, then, are simply the controls that govern the occurrence of the triggering pulse and hence the start of each sweep. These controls are the two sets of concentric knobs located to the upper right of the CRT screen.

(1) Triggering Mode Switch. To suit the type of triggering to the requirements of a particular waveform or range of frequencies, the red five-position switch marked Triggering Mode is provided. The use and range of each mode follows:

DC Mode has a useful range from dc to about 5 or 10 megacycles per second. This Mode accomodates slowly changing waveforms. As taught in "Waveform Analysis" any periodic wave has an average dc level. If this dc level is other than zero the presence of the dc component will influence the triggering. DC Mode permits the dc component to affect the triggering of the sweep.

AC Slow Mode has a useful range from 20 cycles to about 5 or 10 megacycles per second, and differs from the DC Mode in that it filters out the dc component of the observed waveform. In this Mode triggering is independent of the average dc level of the observed waveform.

AC Fast Mode has a useful range from about 10 kilocycles to about 5 or 10 megacycles. It is similar to AC Slow but in addition to filtering out the dc component of the waveform, it also removes any component below 2 kilocycles. This is especially useful in preventing the trigger from encountering interference from hum or other low frequency components.

Automatic Mode is useful for general purpose viewing of waveforms in the frequency range of about 60 cycles to 2 megacycles.

DOCID: 3726612

HF Sync Mode is useful in the range of 5 megacycles to about 25 megacycles. Actually this Mode permits synchronization with signal frequencies as high as 30 megacycles with sweep rates to 200 kilocycles. (This latter sweep rate would cause 150 cycles of a 30 megacycle signal to be displayed.)

(2) Trigger Slope Control. The black knob marked Trigger Slope accomplishes two functions--first, to select the source from which the sweep will be triggered and second, to select the slope for which triggering will occur. The three selections available are from the ac line voltage, from the observed signal, and from some external source. The respective positions are marked "Line," "Int," and "Ext." In any of these three main positions there are two sub-positions marked "+" and "-". The + position causes the trigger to occur during the rising position of the triggering waveform and the - position causes it to occur during the falling position of the triggering waveform.

(3) Stability Control. The red knob labeled stability is the control which determines whether the sweep will free-run or be triggered by one of the signals selected by the Trigger Slope Control. When the Stability Control is in a clockwise position the sweep will free-run and when in a counterclockwise position the sweep is triggered. If the control is turned to the extreme counterclockwise position until it clicks, it engages a condition marked Preset.

Preset is a condition in which the Stability Control will ordinarily be operated for DC, AC Slow, or AC Fast Modes. There are instances when the control may be used in some variable position when using one of these Modes. In any case, when this control is properly set, there will be no sweep in these three Modes unless there is a signal applied to the oscilloscope input.

When HF Sync Mode is used, the Stability Control is utilized as a synchronization control and in Automatic Mode this control is disabled.

(4) Triggering Level Control. Concentric with the Stability Control is the Triggering Level Control. Its function is to determine the voltage level on the input signal at which the

DOCID: 3726612

sweep (and hence the trace) will start. In the Automatic and HF Sync Modes this control is disabled, but it is used with the DC, AC Slow and AC Fast Modes.

For any of the latter three Modes, the control is first set either extreme counterclockwise or extreme clockwise (marked - and + respectively) and then adjusted toward mid-position (zero) until a stable display of the input waveform appears. (For low amplitude input signals triggering can only occur near the zero setting.) The control may then be adjusted further until the trace starts at the desired point on the waveform. Negative settings cause the triggering to occur during the negative level of the input and positive settings cause the triggering to occur during the positive level.

It should be called to the student's attention that in the DC Mode there will be interaction between the Vertical Position Control on the Preamplifier and the Triggering Level Control. This is because the dc component of the waveform is permitted to affect the position of the waveform on the screen and hence will affect the triggering level.

3. Delaying Sweep

Delaying Sweep (Figure 97), as such, is not used by students of this course, and therefore the controls dealing with this section will not be discussed.

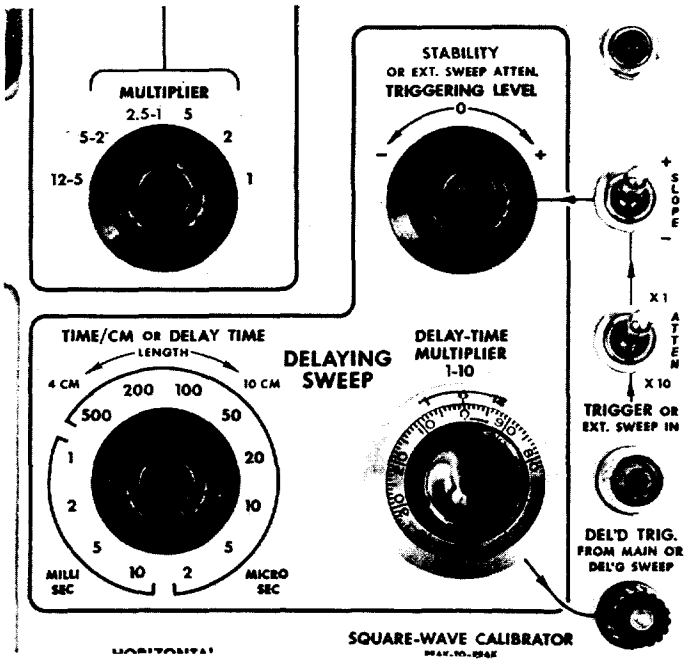


FIGURE 97. DELAYING SWEEP SECTION CONTROLS.

DOCID: 3726612

4. External Sweep

Unlike the controls of Main Sweep and Delaying Sweep, the External Sweep controls shown in Figure 98 are not grouped in a section by themselves. Instead these controls are scattered and have functions other than that of setting the operation of the External Sweep. Primarily, the position of the Horizontal Display Selector will determine the use of these controls. The controls relating to External Sweep may be considered under two categories, namely, those having to do with the Horizontal Deflection Amplifier System and those required to turn on the External Sweep circuits. The former will be discussed under Horizontal Deflection Amplifier System and the latter are discussed here.

To use an external sweep, two basic operations are required. First it is necessary to disable the Main Sweep to prevent interference with the external sweep, and secondly it is necessary to prepare the oscilloscope to handle the external signal. The first requirement is met by (a) and (b) below; the second requirement by (c) and (d).

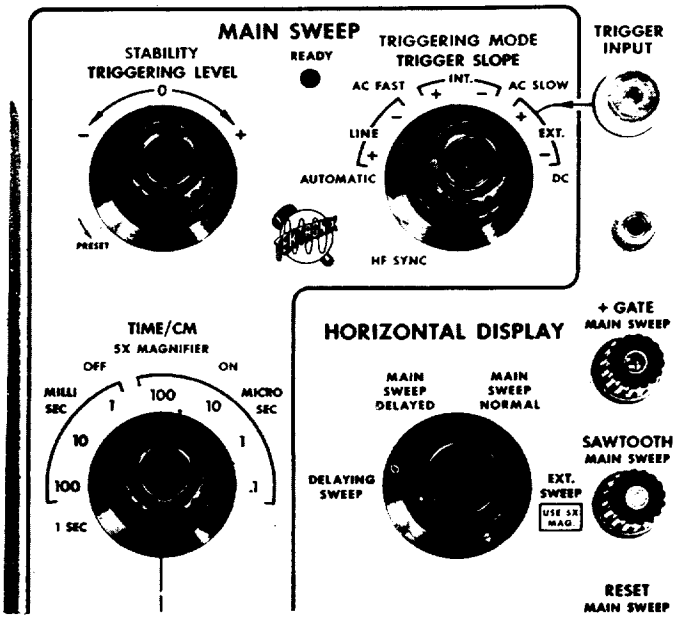


FIGURE 98. CONTROLS TO APPLY EXTERNAL SWEEP.

DOCID: 3726612

(a) Stability Control. It will be remembered that clockwise positions of this control allow free-running operation of the Main Sweep and counterclockwise positions require a trigger. The exception is that if this control is set fully counterclockwise yet not far enough to engage Preset, the main sweep generator will be disabled.

(b) Triggering Mode Switch. The Stability Control is operative for all positions of the triggering mode switch except Automatic. Therefore to make the setting of the stability control effective, the triggering mode switch must be set to some position other than Automatic.

(c) Horizontal Display Selector. Circuit changes in the horizontal deflection amplifier system are necessary in order to apply external sweeps. These changes are accomplished internally by means of the Horizontal Display Selector.

(d) 5X Magnifier. A necessary change when using external sweep is to increase the gain of the horizontal deflection amplifier system. This is done by turning the 5X Magnifier to the ON position. This is a must for proper operation.

C. Horizontal Deflection Amplifier System

The horizontal deflection amplifier system in this precision oscilloscope is more complex than that of the basic oscilloscope just as is the vertical deflection amplifier system. In fact, similar to the vertical deflection amplifier system, the horizontal system also consists of two basic amplifiers but at this point the similarity ends. The two amplifiers in the horizontal system are the Sweep Amplifier and the External Sweep Amplifier. Unlike the vertical deflection system which has all the operator controls in the preamplifier, this system has controls for both portions of the horizontal system. These are shown in Figure 99.

1. Sweep Amplifier

The Sweep Amplifier feeds directly to the horizontal deflection plates of the CRT and is always in use when the oscilloscope is operated regardless of the type or source of sweep or other horizontal signal.

(a) Input. The input to this amplifier is strictly internal. The particular input is selected by the four positions of the Horizontal Display Selector already discussed.

DOCID: 3726612

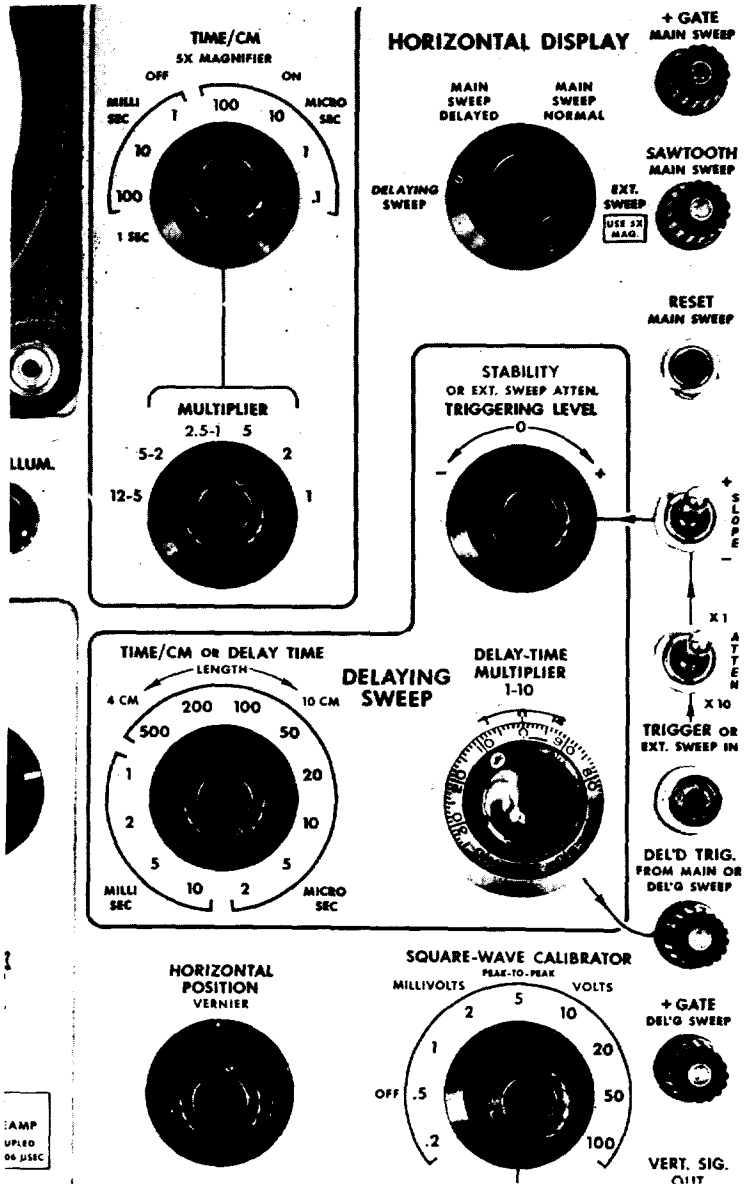


FIGURE 99 HORIZONTAL DEFLECTION AMPLIFIER SYSTEM CONTROLS.

(b) Controls. Only two controls are provided for the operation of the Sweep Amplifier. They are the Horizontal Position Control (with vernier) and the 5X Magnifier.

(1) The Horizontal Position Control. The function of this control is to horizontally center the oscilloscope trace on the screen. It is the black knob located directly above the Power On switch and Pilot Lamp.

DOCID: 3726612

This control has a vernier which is the red knob in the center of the Horizontal Position Control. The vernier has five times the resolution of the coarse control.

(2) The 5X Magnifier. The only gain adjustment for the Sweep Amplifier is the 5X Magnifier. It has two positions only--ON and OFF. It accomplishes two things--to expand the sweep (previously described in Time Controls) and to provide additional gain when external sweep is used.

2. External Sweep Amplifier

One of the oscilloscope's amplifier circuits normally used for another purpose is adapted for use as an auxiliary amplifier when it is desired to place an external signal on the horizontal deflection plates. When adapted to this usage by the Horizontal Display Selector the amplifier becomes the External Sweep Amplifier.

(a) Input. A coaxial connector located on the extreme right of the oscilloscope face and adjacent to the Delaying Sweep controls is provided for inputs to the External Sweep Amplifier. This UHF coaxial connector is marked Trigger or External Sweep In. Signals applied to this input are subject to certain limitations. First of all, the circuits in this section were designed primarily for sawtooth waveforms, hence, application of other waveforms may result in distortion. If sinewaves are used for horizontal input, the highest frequency capable of giving a full 10 centimeter sweep without resultant distortion is 240 kilocycles. Secondly, if the input to this section has a dc component, the dc balance of the horizontal deflection system will be upset.

(b) Controls. There are three controls available to the External Sweep Amplifier. These are located above the input connector in the sequence Attenuator Switch, Slope Switch, and the External Sweep Attenuator.

(1) Attenuator Switch. The Attenuator Switch has two positions, X1 and X10. When necessary, it is used to reduce the external sweep amplitude by a factor of 10. If the external sweep is greater than 10 volts the X10 position is used, and if less than 10 volts the X1 position is used.

DOCID: 3726612

(2) Slope Switch. Sometimes it is desirable or necessary to invert the polarity of the external sweep. This inversion is easily accomplished by positioning the Slope Switch.

(3) External Sweep Attenuator. This is the control in the delaying sweep section labeled stability or External Sweep Attenuator. In its function with the External Sweep Amplifier it serves as a continuously variable gain control to adjust the sweep length as desired. The continuous control of amplifier gain afforded here in addition to the 10:1 fixed attenuator Switch provides a 100:1 adjustment of horizontal deflection sensitivity.

D. CRT Controls and Accessories

Although there are certain refinements for improved performance, the cathode ray tube in this equipment does not differ in principle from those discussed in the preliminary portion of this study. There is little, therefore, to be said

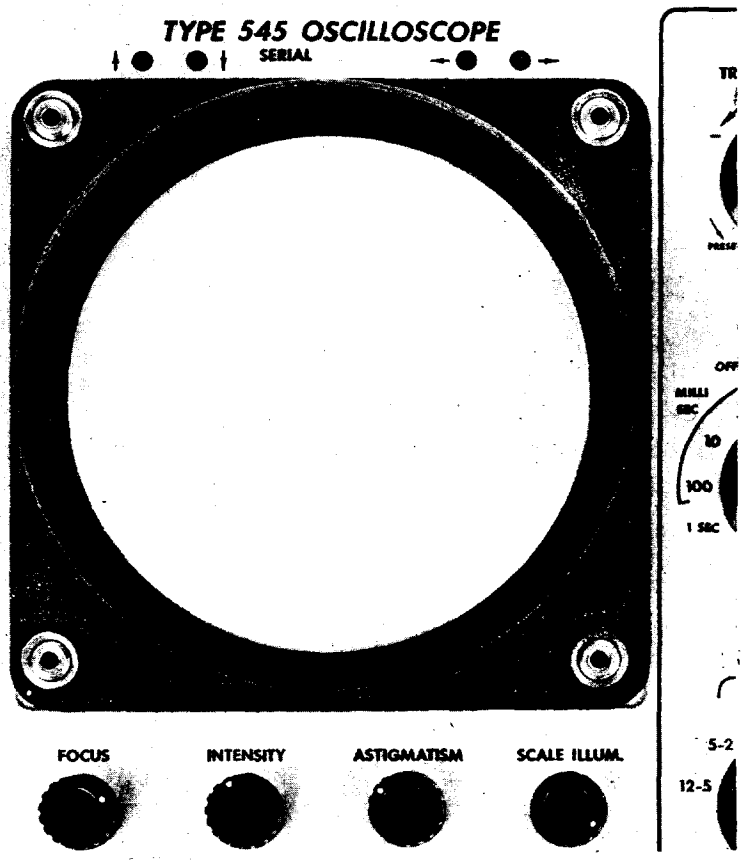


FIGURE 100. CRT CONTROLS AND ACCESSORIES.

concerning the CRT itself but somewhat can be said of its associated controls and accessories which are illustrated in Figure 100. In fact, apart from those having to do with auxiliary functions, the only controls and indicators remaining to be discussed are those relating directly to the CRT.

1. Filter and Graticule

It is not apparent in Figure 100, but directly in front of the CRT screen is a green optical filter designed to provide a maximum trace contrast in ordinary room lighting. This filter must be used for photography to eliminate reflections.

In front of this is a pane of transparent plastic engraved on one side with grid lines accurately spaced one centimeter apart and tick-marked at one-fifth centimeter (two-tenths of a centimeter) intervals. This plastic pane with its engraved scale is called a graticule. The optical filter and graticule are held in place by a bezel and four mounting nuts which are only hand-tight and easily removable.

The graticule scale (which does show in the illustration) measures four centimeters by ten centimeters. The scale is used in conjunction with the volts/cm switch to read peak-to-peak voltage, and with the time/cm control to read sweep time. The reading in either of these usages is simply applied dimensional analysis where the calibrated setting of the control is multiplied by the number of centimeters of deflection observed on the graticule. For example:

$$\text{volts} = \frac{\text{volts}}{\cancel{\text{centimeters}}} \times \cancel{\text{centimeters}}$$

$$\text{sweep time} = \frac{\text{time}}{\cancel{\text{centimeters}}} \times \cancel{\text{centimeters}}$$

2. Scale Illumination

For proper illumination of the graticule, the optical filter must be mounted next to the cathode ray tube followed by the graticule placed with its engraved side next to the filter.

The illumination of the graticule may be either of two colors: red or white. The choice of color is determined simply by rotating the graticule 180 degrees. The white is used whenever it is desired to

photograph the scale along with a waveform, whereas the red illumination will not photograph at all.

Very closely associated with the graticule is the Scale Illumination control. This control is used to adjust the intensity of the lights that illuminate the graticule scale to suit the room lighting conditions.

3. Intensity and Focus

The purpose and operation of the Intensity Control and Focus Control on this oscilloscope do not differ from those of the basic oscilloscope, therefore no further explanation will be offered here. However, due to the damage that can result from careless usage it must be emphasized that the electron beam in the CRT has a very high energy imparted to it by the 10,000 volts accelerating voltage. This energy, when given up to the screen, produces a spot intensity that is capable of destroying the screen material if the beam is allowed to remain stationary. Therefore, DO NOT LEAVE A FIXED BRIGHT SPOT OR A SLOW SWEEP ON THE SCREEN. As a precaution the beam should either be defocused or the intensity turned all the way counterclockwise whenever the instrument is unattended.

4. Astigmatism Control

The function of the Astigmatism Control is to correct the defocusing that occurs at the extremities of the oscillograph trace due to the difference in distance travelled by the electron beam. The Astigmatism Control is adjusted in conjunction with the Focus Control to obtain a sharp, clearly defined trace.

5. Beam Position Indicator

Observe in Figure 100 that there are two pairs of neon lamps located immediately above the CRT bezel. These are unlabeled and marked only with an arrow pointing up, down, left, or right. The lamps with their arrows are Beam Position Indicators used to indicate the position of the beam with respect to the center of the graticule. Used in conjunction with the Vertical Position Control on the preamplifier and the Horizontal Position Control on the main unit, the Beam Position Indicators make it possible to center the beam in the graticule without turning the intensity above visual cutoff.

To illustrate the operation, when the beam is off the graticule in an up-down direction, either a lamp beside the arrow pointing up or a

DOCID: 3726612

lamp beside the arrow pointing down will be lit indicating the way the beam is off the graticule. The Vertical Position Control is then adjusted to a position where the lamp is extinguished without the other lamp of the pair becoming lit. Similarly, when the beam is off the graticule in a left-right direction either a lamp beside the arrow pointing left or one beside the arrow pointing right will be lit. The Horizontal Position Control is then adjusted so that neither of the lamps of this pair is lit.

E. The Square Wave Calibrator

An auxiliary function contained in the 545 is the Square Wave Calibrator. This is not a part of the oscilloscope circuits but is an auxiliary equipment contained within the unit to provide a signal for checking the calibration of the vertical deflection system and for calibrating the oscilloscope probes to insure accuracy in pulse and transient measurements. A set of concentric red and black knobs (shown in Figure 101) located below the Delaying Sweep section are used to turn on the calibrator and to set the amplitude of its output. The amplitude may be selected by the black control in steps from 2 to 100 millivolts when the red control is in the millivolt position, and from 2 to 100 volts when the red control is in the volt position.

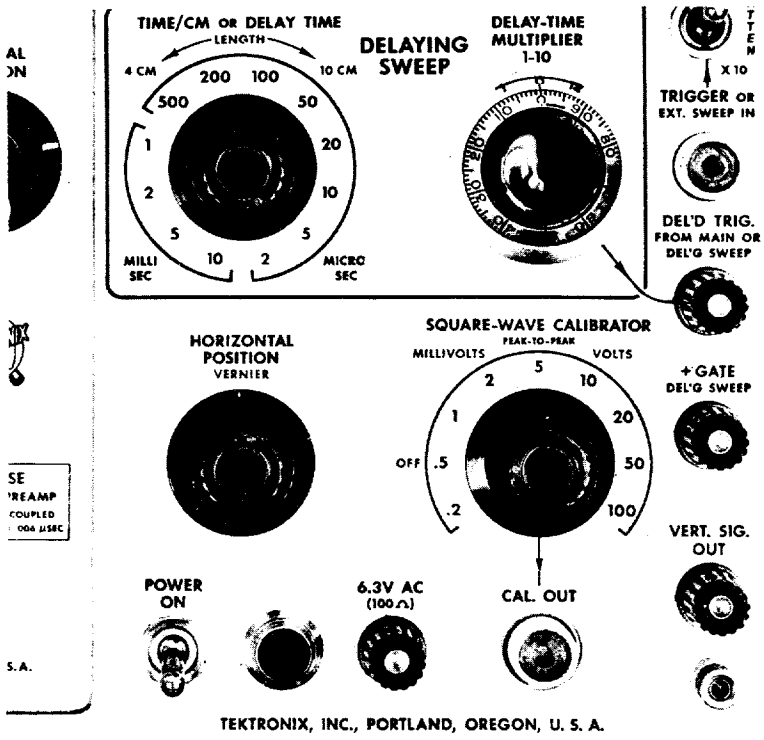


FIGURE 101. SQUAREWAVE CALIBRATOR.

DOCID: 3726612

The output of the calibrator is a squarewave whose fundamental frequency component is approximately one kilocycle and is available at the UHF coaxial connector labeled Cal. Out.

IV. OPERATING INSTRUCTIONS

Whenever operating any oscilloscope for the first time, it is advisable to obtain a trace and a waveform by the simplest and easiest means possible. Then after acquiring an understanding of the controls and their operation it will become a fairly simple matter to readjust to meet a specific operating or waveform requirement. This general principle is no less applicable to operating the laboratory oscilloscope used in this course.

It is the intention here to follow this suggested procedure by first giving general instruction on obtaining a trace and a waveform and then the specific changes necessary to achieve the desired result.

A. General Procedure

The present objective is to indicate what must be done with a particular control or controls to obtain a desired result. The purpose of each of the controls has already been described and will not be repeated. Therefore, if the student desires to know why he is making an adjustment it is recommended that he review the earlier explanation. The procedure outlined below presupposes that the oscilloscope has been equipped with a type K plug-in preamplifier.

1. Preliminary Information

Before turning ON the oscilloscope several operating characteristics which have not been explained elsewhere should be understood.

(a) Time Delay Relay. To allow adequate tube warm-up before the high voltage comes on, the instrument is equipped with a time delay relay. This provides a time delay of approximately 25 seconds. Should the power be turned off or the line cord become disconnected (even if for just an instant) while the oscilloscope is in operation there will be a time delay of 25 seconds before the unit can be restored to operation. The completion of the time delay is evidenced by a ready light located in the top center of the Main Sweep Section.

(b) Thermal Cut-Out. The oscilloscope is kept cool by filtered forced air. The air intake is located at the rear of the unit.

If the intake should become obstructed by any means such as a dirty filter or by having the unit backed up against a wall, the instrument will overheat and actuate a thermal cut-out. This automatically turns the oscilloscope off until the internal temperature has again dropped to a safe operating level.

(c) Probes. Readers of this text will use wide-band or fast rise time preamplifiers only. Certain probes are not applicable to this type unit. In particular, the type P510A probe should not be used with these preamplifiers because the probe tends to ring at about 50 megacycles. A fast rise time or wide band preamplifier permits this ringing to be displayed on the CRT screen superimposed on the desired waveform. Probes of the P400 series (such as the P410 supplied with the oscilloscope) may be used with these preamplifiers without undesirable consequence.

P400 series probes with their color codes and the amount by which they attenuate the input signal are tabulated below.

PROBE TYPE	NOSE COLOR	ATTENUATION RATIO
P405	Green	5:1
P410	Brown	10:1
P420	Red	20:1
P450	Clear, Green inside	50:1
P450-L	Clear, Green inside	50:1
P4100	Clear	100:1

2. Preparatory Operation

Preparatory operation is a prescribed procedure to ready the oscilloscope for use and to check the calibration of the probe. A simple trace is first obtained; then the output of the Square Wave Calibrator is displayed so that the probe adjustments can be made.

Before plugging in the power cord be sure the power switch is OFF. The power cord should then be connected to a 117 volt 60 cycle source. Before plugging in or removing any preamplifier the power switch should always be OFF. (For our use the oscilloscope will be equipped with a Type K plug-in preamplifier and a Type P410 probe.) Front panel controls are then set as follows:

DOCID: 3726612

Intensity	Fully CCW
Focus	Centered
Astigmatism	Centered
Power Switch	ON
Horizontal Display	Main Sweep Normal
Horizontal Position	Centered
Vertical Position	Centered
Calibrator	Black, 10; Red, volts
Time/cm	1 millisecond
Multiplier	1
5X Magnifier	OFF
Triggering Mode	Automatic
Trigger Slope	+ INT
Stability	any
Triggering Level	any

After setting the controls as indicated above, the Intensity should now be adjusted for useful brightness and the Focus and Astigmatism adjusted for the sharpest possible trace. Then a coaxial cable with a P410 probe should be connected to the Input connector of the Type K preamplifier. Insert the probe tip into the CAL. OUT connector. The preamplifier controls must be set as follows:

AC-DC Switch	AC
Vertical Position	Centered
Volts/cm	.5
Variable Gain	Calibrated (fully CW)

Final adjustments of Focus, Intensity and Astigmatism may now be made to obtain optimum sharpness and brightness. The waveform should be centered in the graticule scale by means of the Vertical Position and Horizontal Position controls. If these instructions have been followed closely, a waveform consisting of 7-13 rectangular pulses 2 centimeters in amplitude should be displayed across the screen.

3. Probe Calibration

The waveform observed during the procedure just outlined although essentially square, may have either a spike or a curve on its leading edge like those illustrated in Figs. 102A and 102B. If so, it is because the probe is out of calibration. If accurate pulse and transient measurements are to be made this distortion must be eliminated. Even when

DOCID: 3726612

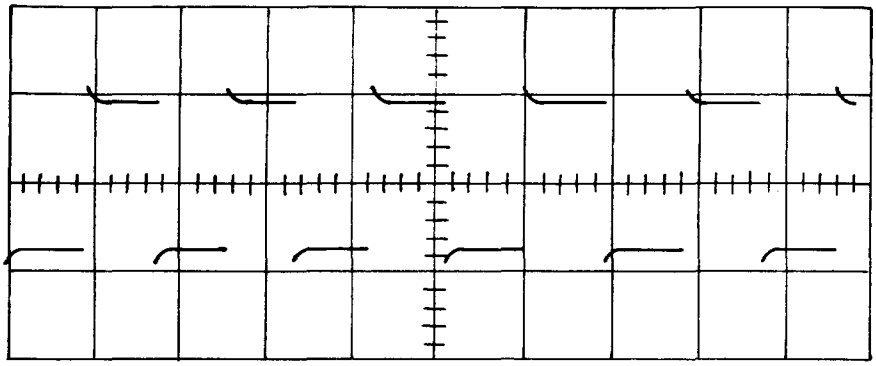


FIGURE 102a. OVER COMPENSATED.

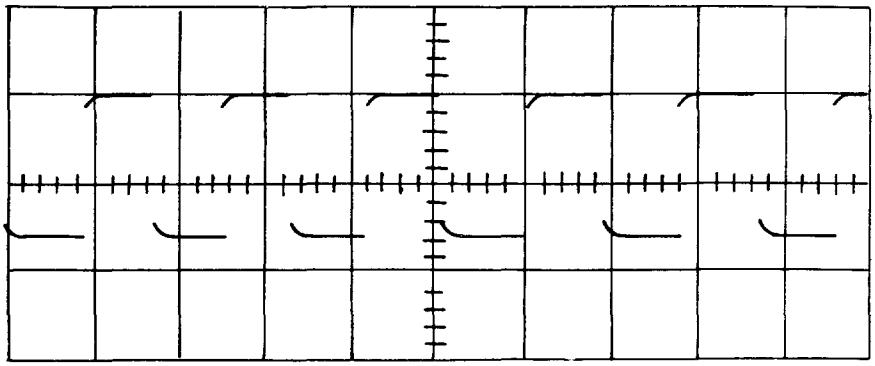


FIGURE 102b. UNDER COMPENSATED.

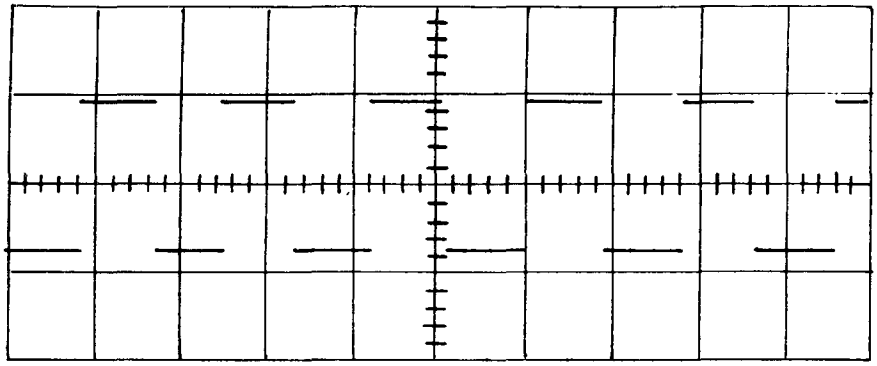


FIGURE 102c. CORRECTLY COMPENSATED.

FIGURE 202. EFFECTS OF PROBE ADJUSTMENT.

only sinewaves are to be observed it is absolutely necessary that the probe be calibrated if amplitude measurements are to be accurate.

The distortion is created at the input of the preamplifier when the input capacitance of the probe is not adjusted so as to balance out the effects of the input capacitance of the preamplifier. To compensate for variations in input capacitance from one preamplifier to another an adjustable capacitor is placed in the body of the probe. The distortion is easily removed by varying the screw adjustment in the body of the probe until a good flat-top square wave like that shown in Figure 102C is obtained. The probe is then said to be calibrated.

B. Measurement of Wave Period

It has been taught that in order to observe waveforms it is necessary to apply an extremely linear sweep voltage to the horizontal deflection plates of the CRT and to apply the signal to be studied to the vertical deflection plates. In the 545, positioning the Horizontal Display Switch to Main Sweep Normal supplies the required linear sweep to the CRT and the signal to be observed is applied by means of the input to the preamplifier.

Main Sweep Normal operation is used for all general purpose viewing of waveforms and for wave period measurements. This main sweep operation is used in the Automatic Triggering Mode, in the general procedure just given, but for other applications it can be used in any of the other modes available. The general settings are shown in the table below.

MAIN SWEEP NORMAL

Modes	Trigger Slope	Triggering Level	Stability
Automatic	± Internal	Any (Not used)	Any (Not used)
DC	± Internal	Fully CW or CCW	PRESET
AC Slow	± Internal	Fully CW or CCW	PRESET
AC Fast	± Internal	Fully CW or CCW	PRESET

The following measurement will further illustrate the use of Main Sweep Normal and will demonstrate how a precision sweep can be used for measuring wave periods. For this measurement a mode other than automatic has been selected

DOCID: 3726612

so the triggering level can be selected to initiate the trace at any desired point on the waveform. AC Slow has been chosen to the exclusion of DC mode because the dc component (if any) is of no interest to us at present, and it has been preferred over AC Fast so that low frequency signals may be measured. The procedure should be put into practice as soon as possible so the student can develop a "feel" for the oscilloscope and its controls.

1. Preparation of the Equipment

After the oscilloscope has been put into operation and the probe calibrated as outlined in the general procedure, the oscilloscope and the associated equipment should be adjusted as follows:

(a) V/cm control ---------- .5

(b) Variable gain ---------- Calibrated

(c) A sinewave output from a Heathkit Sine-Square generator (Model AG-10) is patched into the oscilloscope preamplifier through a P410 probe.

(d) The generator amplitude control is adjusted for convenient waveform amplitude.

(e) The Horizontal and Vertical Position controls are adjusted until the waveform is centered in the graticule.

(f) It is ascertained that the Main Sweep controls are set as follows:

Triggering Mode ----------AC Slow
Trigger Slope ---------- + Int.
Triggering Level ----------Fully CW or CCW
Stability ----------Preset

(g) The Triggering Level control is adjusted toward "0" until the waveform reappears and begins exactly on the horizontal center line of the graticule.

2. Preparation of the Display for Measurement

(a) While maintaining good centering, the generator amplitude control is adjusted until the total vertical deflection is 4 centimeters.

DOCID: 3726612

(b) The Time/cm control and Multiplier (black scale only) are adjusted for a display of 1 to 3 cycles.

(c) The Volts/cm switch is positioned to .05 to cause the waveform to cross the horizontal center line of the graticule almost vertically. This will also drive the top and bottom of the waveform off the screen. The presentation on the oscilloscope screen should be similar to Figure 103. The abrupt crossing of the center line is necessary in order to accurately measure the time between crossovers.

(d) After the last adjustment it may be necessary to re-position Intensity, Focus, and Astigmatism for a sharp display.

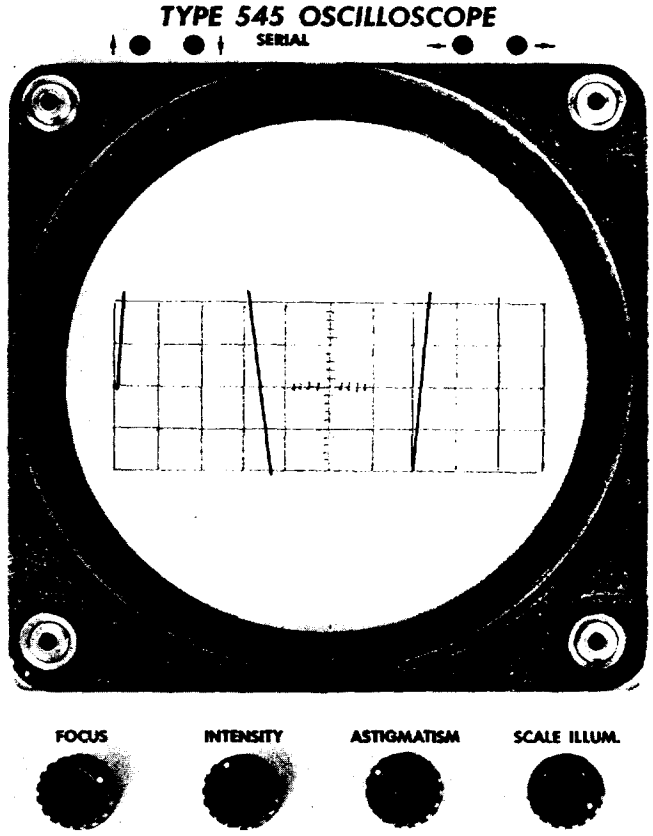


FIGURE 103. WAVEFORM AS IT SHOULD APPEAR IN STEP 2(C).

DOCID: 3726612

3. Wave Period Calculation

(a) The distance along the horizontal center line of the graticule from the beginning of the waveform to the second cross-over represents the period of the wave (Figure 103).

(b) The wave period is given by the product of the distance measured on the graticule, the time/cm control setting, and the Multiplier setting (black scale only) thus, --

$$\text{Period} = \text{cm} \times \frac{\text{time}}{\text{cm}} \times \text{multiplier}.$$

4. Evaluation of the Procedure

When the student is becoming oriented in the procedure for making wave period measurements, he should make the following analysis.

(a) Calculate the wave period.

(b) Without changing control settings on either the oscilloscope or signal generator, move the probe from the sinewave output to the squarewave output.

(c) Adjust the signal generator amplitude control for maximum vertical deflection.

(d) Calculate the period of the square wave.

(e) Are the periods found in (a) and (d) the same? Why?

(f) Calculate the frequency represented by the period.

(g) What does (f) tell us about two waves that have the same period?

(h) If the calculated frequency does not agree with the signal generator dial reading, note the size of the error.

(i) Which do you consider the more nearly correct frequency? Why?

(j) Repeat the procedure and its analysis with other arbitrary signal frequencies until it becomes natural to perform.

DOCID: 3726612

(k) Adapt the procedure to an unknown signal from an unknown source. The usefulness of this method of period measurement is extensive in that it can measure the period of any periodic wave regardless of its shape. From this the fundamental frequency of the wave is readily computed. The main limitation in the general method employed is that an oscilloscope with an accurately calibrated sweep (such as in the 545) is required.

C. Measurement of Wave Frequency

A rapid and accurate method of measuring the frequency of a sinewave is by means of Lissajous figures. (One such figure was shown in Figure 91.) Making frequency measurements by means of Lissajous figures will demonstrate how to operate the oscilloscope using External Sweep but before explaining the oscilloscope operation it is first necessary to explain the development of Lissajous patterns.

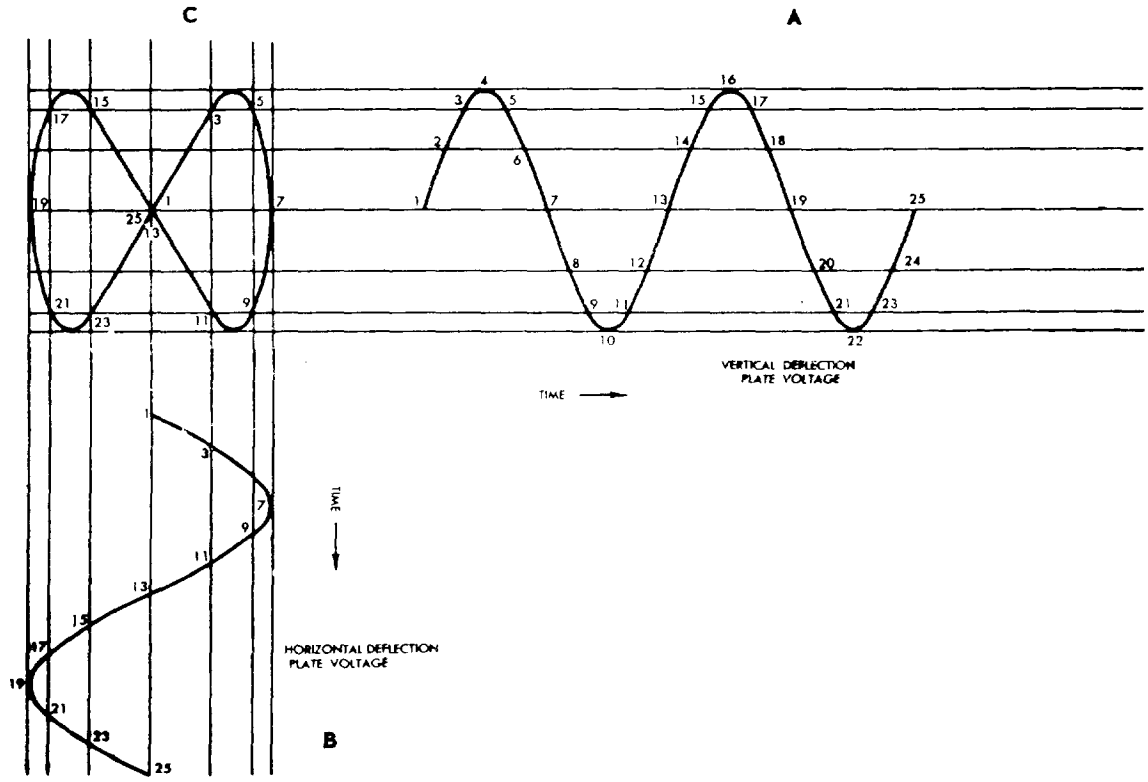


FIGURE 104. DEVELOPMENT OF A LISSAJOUS PATTERN.

DOCID: 3726612

1. Development of Lissajous Figures

Lissajous patterns are developed whenever the signal applied to the vertical deflection plates and the sweep voltage applied to the horizontal deflection plates are both sinewaves. For the creation of a true Lissajous figure the applied signals should both have the same amplitude. The resulting pattern is a figure that is characteristic of the difference in frequency between the two signals. Because of this characteristic these patterns find their most important application in determining the frequency of one of the signals when the frequency of the other is known.

The development of a Lissajous figure is illustrated in Figure 104. Sinewave A is applied to the vertical deflection plates in twice the frequency of sinewave B applied to the horizontal deflection plates. Note that in the illustration both A and B are initially in phase. Both waves have been divided into equal time intervals which have been numbered from 1 to 25. This relates the two waves as to time. The combined displacement of the electron beam as produced by each wave at the same instant will result in the oscilloscope image, C, which is a Lissajous figure.

Lissajous figures form many different patterns determined by both frequency and phase. Some of these are shown in Figure 105.

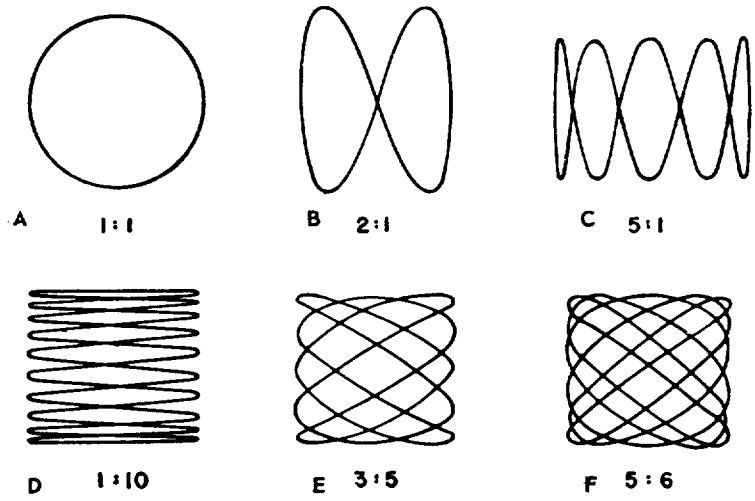


FIGURE 105. LISSAJOUS PATTERNS FOR VARIOUS FREQUENCY RATIOS.

DOCID: 3726612

To make use of this quick, accurate method of measurement it is necessary to know how to interpret the pattern. Although some of these patterns appear complex, interpreting them for the frequency ratio they represent is not at all difficult. It is only necessary to draw or imagine a horizontal line tangent to the top of the pattern and a vertical line tangent to one side of the pattern. An example is given in Figure 106A.

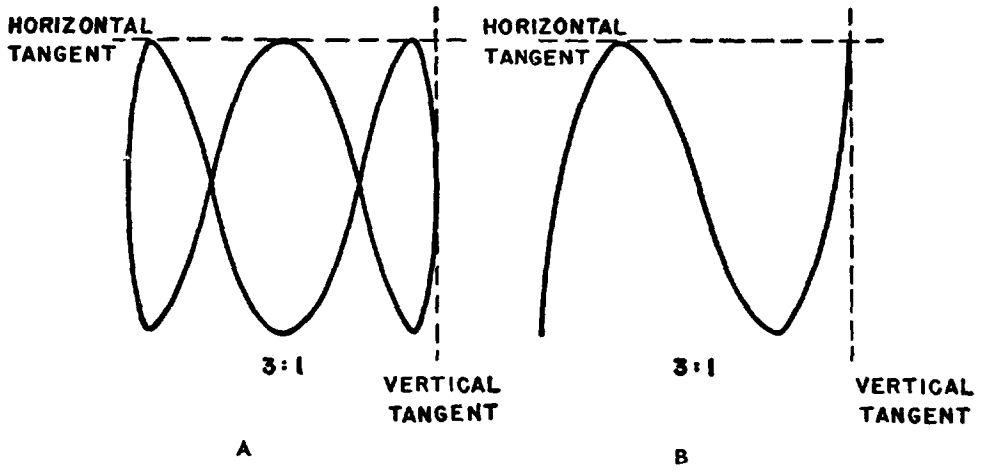


FIGURE 106. INTERPRETING FREQUENCY RATIOS.

The ratio of the frequencies of the waves composing the Lissajous figure is found by counting the number of points, T_H, where the pattern is tangent to (touches) the horizontal line, and dividing this by the number of points, T_V, where the pattern is tangent to the vertical line. The ratio of the frequency of the signal applied to the vertical deflection plates, F_V, to the frequency of the signal applied to the horizontal deflection plates, F_H, is equal to the ratio of the number of points tangent to the horizontal line, T_H, to the number of points tangent to the vertical line, T_V. Expressed as an equation this is:

$$\frac{F_V}{F_H} = \frac{T_H}{T_V}$$

This ratio is very important and should be memorized. It is recommended that the student practice using this relationship by verifying the ratios of the patterns shown in Figures 105 and 106B.

DOCID: 3726612

(a) Phase Effects. Figures 106A and 106B do not look alike although they both express the same frequency ratio of 3:1. The difference in appearance is due to a difference in phase only. As an example of the effect that phase difference can have on a figure, Fig. 107 shows some of the possible patterns that can exist when there is a 1:1 frequency ratio but different phase relationships.



0°	22.5°	45°	67.5°	90°	112.5°	135°	157.5°	180°
360°	337.5°	315°	292.5°	270°	247.5°	225°	202.5°	

FIGURE 107. EFFECTS OF PHASE DIFFERENCE.

A circle results from a 90° phase difference, a straight line from either an in-phase condition or a 180° phase difference, and elliptical patterns for all intermediate phase differences. It is not intended that students of this course be able to measure phase by means of Lissajous figures, therefore it is not necessary that the student concern himself with the matter of phase effects beyond the fact that phase can affect the appearance of the pattern but does not affect the ratio of T_H/T_V.

(b) Determining Unknown Frequency. It should be obvious that a Lissajous pattern does not give a frequency read-out. It only provides a very accurate ratio of two frequencies. If one of the two frequencies is known (this is generally the signal on the horizontal deflection plates) it is possible to find the unknown frequency from the known frequency and the ratio. Using dimensional analysis:

$$\frac{F_V}{F_H} \times F_H = F_V$$

If, for example, the horizontal frequency used in Figure 106 had been 200 cycles, we could substitute the known value into the above relationship and find the vertical signal to be 600 cycles, thus

$$\frac{3}{1} \times 200 = 600$$

DOCID: 3726612

The frequency ratio from a Lissajous pattern is accurate. The calculated frequency is only as accurate as the signal generator that provides the reference signal and the operator's ability to use it. This necessitates the use of a stable and accurately calibrated signal generator and an operator whose methods are not slip-shod. The only limitations on the Lissajous method of frequency measurement are that it can be used only with sinewaves, and their use is impractical for frequency ratios greater than 10:1 or 1:10. The last fact is in evidence from the pattern in the lower left corner of Figure 105.

2. Preparing the Equipment

For most oscilloscope viewing the sawtooth voltage from the sweep generator provides the time-base for the signal being observed or measured. To obtain Lissajous figures, however, it is necessary to use a sinewave as the time-base. This sinewave must be obtained from an external source. To prepare the oscilloscope for this type operation the Main Sweep Generator must be disabled and the Horizontal Deflection Amplifier system must be set up to handle the external time-base. After the oscilloscope has been put into operation and the probe calibrated as outlined in the general procedure the oscilloscope and associated equipment are adjusted as follows:

(a) Oscilloscope settings -
Stability ---------- fully CCW (but not Preset)
Triggering Mode---------- any except automatic
Horizontal Display--------- EXT Sweep
5X Magnifier---------- ON
X1, X10 attenuator--------- X1
Stability or EXT Sweep Attn.
(In Delaying Sweep Section)------45° from Full CCW
Volts/cm control ---------- .1

(b) The sinewave output from a Heathkit Sine-square generator (Model AG-10) is patched into the oscilloscope preamplifier through a P410 probe (vertical input).

(c) The frequency setting is arbitrary.

(d) Without use of a probe, a sinewave from a Hewlett Packard 200 CD wide range oscillator is patched into the Trigger or Ext. Sweep In connector (horizontal input).

DOCID: 3726612

(e) The frequency setting is arbitrary.

(f) The Vertical and Horizontal Position controls are adjusted until the waveform is centered in the graticule.

(g) The amplitude controls of the signal generators are adjusted for about 4 centimeters deflection both horizontally and vertically.

3. Preparing the Display for Measurement

The signal applied to the preamplifier (vertical input) will be considered as the unknown frequency. The signal applied to the Ext. Sweep In connector (horizontal input) will be considered as the reference frequency. Whenever the relationship between these two frequencies is not a ratio of two integers, the Lissajous pattern will seem to revolve as if it were a hollow, transparent cylinder. The rate at which the pattern revolves will vary as the frequency relationship is varied. To prepare the display for measurement the following steps are taken.

(a) The frequency of the reference generator is varied until the pattern becomes stationary. This condition indicates that the reference signal is a multiple (or submultiple) of the unknown frequency.

(b) The adjustment is continued (if necessary) until the ratio of tangent points is somewhere between 1:1 and 10:1.

4. Sinewave Frequency Calculation

(a) The ratio of points tangent to a horizontal line to the points tangent to a vertical line is equal to the ratio of the vertical frequency to the horizontal frequency:

$$\frac{T_H}{T_V} = \frac{F_V}{F_H}$$

(b) The unknown frequency is found by multiplying the ratio by the reference frequency:

$$\frac{F_V}{F_H} \times F_H = F_V$$

DOCID: 3726612

5. Evaluation of the Procedure

The student can best gain proficiency in the use of the procedure by practice and by answering questions that emphasize the important points. The following are recommended.

(a) With the reference signal fixed at about 1500 cycles per second, vary the vertical signal to obtain all ratios from 10:1 to 1:10.

(b) Calculate the frequency and compare the calculated frequency to the reading on the vertical signal generator.

(c) By how many cycles do the calculations and the readings differ?

(d) Which signal source do you think might be in error?

(e) If the reference signal generator dial were inaccurate what would happen to the calculation?

(f) Arbitrarily set the vertical source to some frequency without observing the particular dial setting. Determine the frequency just from the horizontal source and the Lissajous ratio. Try this with a number of different dial settings.

(g) Adapt your procedure to an unknown signal from an unknown source.

D. Oscilloscope Photography

Oscilloscope photographs are often called oscillograms. Oscillograms are desirable when it is necessary to make a permanent record of certain waveforms. Also, there are some waveform measurements which cannot be accomplished without the use of photographs. In the 545 provision has been made so that clean, clear photographs can be assured. This is easily accomplished by means of single trace photography. Single trace photography removes the risk of blurring the photograph by successive traces that may not exactly match one another. Main Sweep Delayed operation provides the means by which the main sweep is held inoperative until the operator actuates it and then it sweeps the electron beam across the screen only once. It cannot produce another trace until the operator again actuates the sweep.

DOCID: 3726612

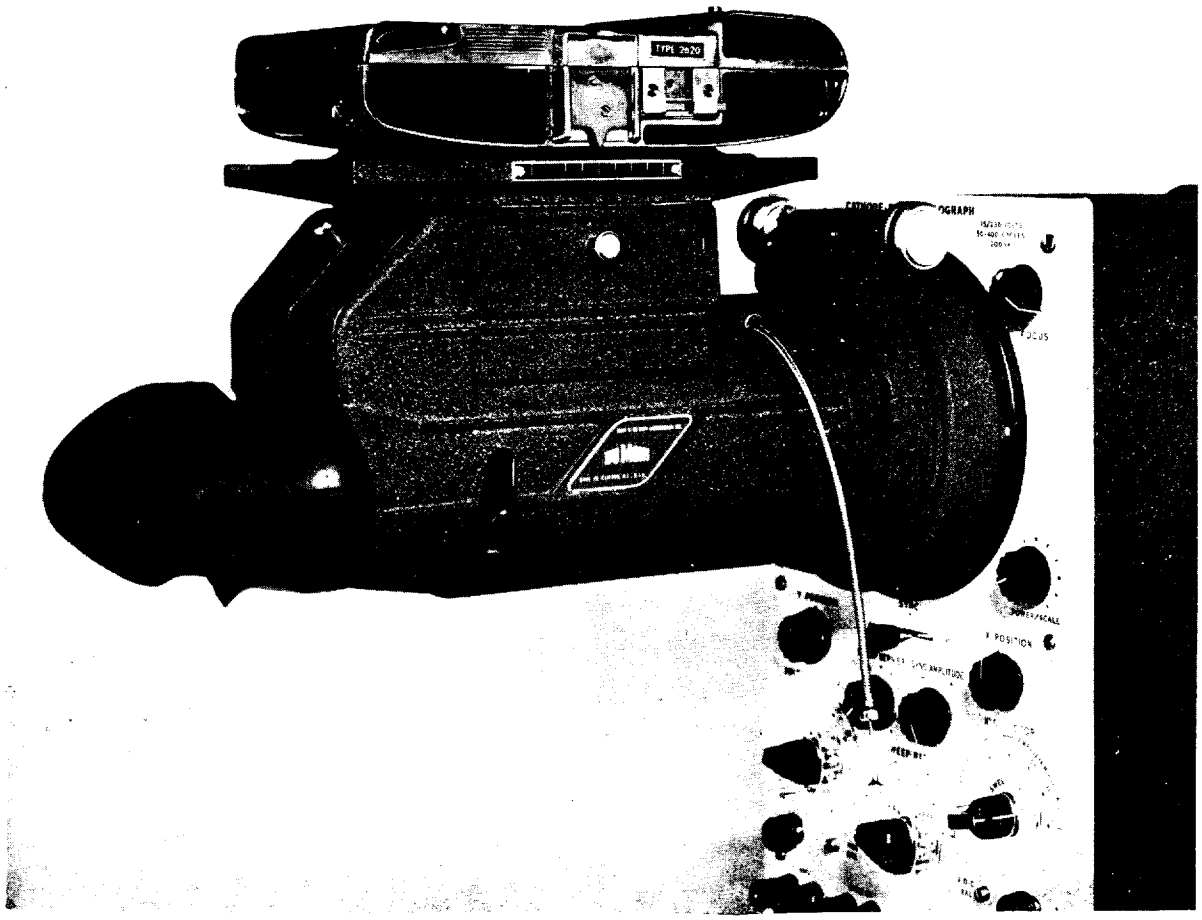


FIGURE 108. DUMONT TYPE 302 OSCILLOGRAPH-RECORD CAMERA.

1. Preparation of the Equipment

It is beyond the scope of this writing to describe in detail the operation of a particular type of oscilloscope recording camera since there are several excellent varieties available. Fig. 108 illustrates the DuMont Type 302 camera which is currently used in this course. However, just as there is a general procedure for taking entertainment photos for the family album, even though a particular model camera may have characteristics requiring a special technique, so a general procedure can be described here.

(a) Mounting the Camera. In order to mount photographic equipment on the 545 oscilloscope it is necessary to remove the four hand-tight knurled nuts that secure the bezel and graticule and place over the original bezel a BE510 camera mounting bezel.

DOCID: 3726612

The BE510 bezel is
Next, the adapter clamp
is secured to the oscillograph camera
the clamp screws. The camera mounting assembly
inserted into the adapter clamp, pressed firmly into place and
positioned. The assembly is then secured by means of the adapter clamp ring screws.

(b) Camera Instructions. Specific instructions for the type camera used should be consulted to determine the requirements of a given type. Apart from the specific instructions there are general requirements as to the shutter adjustment and the adjustment of the aperture in the iris diaphragm. For optimum results the aperture should be set to an f-stop somewhere between f/5.6 and f/8, as found by trial and error for each given sweep speed. The shutter adjustment should be set for "Bulb." This allows the operator to open the shutter just prior to initiating the oscilloscope trace and to close it as soon as the trace is complete.

(c) Oscilloscope Preparation. The oscilloscope and the probe should be checked out according to the general procedure. After the camera assembly is mounted, observe the oscillograph presentation through the viewing door provided in the camera mounting assembly. The oscilloscope adjustments are as follows:

(1) The desired waveform is patched into the preamplifier through a P410 probe.

(2) The volt/cm control is set to some value that will give a presentation having the amplitude that it should have in the photograph.

(3) The horizontal and vertical positioning controls are adjusted to center the waveform in the graticule.

(4) The time/cm control and multiplier are set for a sweep rate that is suited to the desired appearance of the waveform to be photographed.

(5) The Horizontal Display Selector is switched to Main Sweep Delayed and the sweep controls are set as follows:

Triggering Mode----------AC Slow
Trigger Slope ----------any (not used)

DOCID: 3726612

Triggering Level----------Fully cw or ccw
Stability (Main Sweep) ----------Fully cw
Stability (Delaying Sweep)----------Fully cw

When these settings are completed, the pattern traced on the CRT screen will disappear.

(6) The intensity is turned fully cw and then backed off about 10 degrees.

(7) Observing through the viewing door, the operation may be checked by momentarily depressing the RESET button located below and to the right of the Horizontal Display Selector. A single trace should appear on the screen.

(8) If a "halo" appears around the trace in step (2) the intensity should be backed off so that this halo will not appear.

2. Trial Photographs

Trial photographs are an important feature of oscilloscope photography. The primary purpose of the trial shot is to determine proper scale illumination and proper f-stop settings. Making the shot is a simple procedure and accomplished as follows:

(a) Re-check the display to ascertain that desirable results are obtained when the Reset button is depressed.

(b) Set scale illumination to about 50% full scale.

(c) Assuming that the camera is properly loaded with film and ready to take the first photograph, close the viewing door and any other openings that are not normally open during picture taking.

(d) When the camera equipment is ready, open the shutter and hold it open.

(e) Depress the Reset button on the oscilloscope. This starts the trace and extinguishes the "Ready" light at the top of the Main Sweep Section. Close the shutter.

(f) The operation is complete and now it is only necessary for the photograph to be developed.

DOCID: 3726612

3. Evaluation of the Photograph

The trial photographs should enable the operator to make any necessary correction in the f-stop and scale illumination settings.

(a) If the waveform seems over-exposed the f-stop should be increased.

(b) If the waveform seems under-exposed the f-stop should be decreased.

(c) When the f-stop is increased or decreased, the scale illumination should be increased or decreased accordingly.

(d) After a successful trial photograph the equipment is ready to make photographs for record or measurement.

(e) It is very important that the volts/cm control setting, the time/cm control setting and the multiplier setting be written on the photograph if the picture is to serve any useful purpose. Otherwise the effort in making the photographic record is useless.

CONCLUSION:

An oscilloscope is simply a cathode ray tube to which auxiliary equipment has been added so as to improve its performance and to make it more versatile. The oscilloscope is not a difficult instrument if cathode ray tube operation is understood since auxiliary circuits and controls are only a means of regulating the operation of the CRT. Therefore, if the student understands cathode ray tube operation he can understand the operation of a basic oscilloscope. When the principles and operation of the basic oscilloscope are mastered the student can, with a minimum of reorientation, operate any oscilloscope regardless of type.

The theory and examples presented here should be put into practice as soon and as often as possible. In return, the oscilloscope will prove itself an invaluable tool for observation and measurement of waveforms.

DOCID: 3726612

V. PRACTICAL EXERCISES

1. The movement of the beam with no signal applied may be controlled by:

   a. the vertical position control
   b. the vertical gain control
   c. the intensity control
   d. the focus control

2. An oscilloscope can be used to

   a. measure an unknown frequency
   b. measure short time intervals
   c. present a visual picture
   d. all of the above

3. Why should a sharply focused line or spot not be allowed to remain on the screen?

4. A circular Lissajous pattern denotes a frequency ratio of

   a. 2:1
   b. 1:1
   c. 2:3
   d. 10:1
   e. 1:10

5. The pattern appearing on the screen is very dim. To increase the brightness the operator should change

   a. the focus control
   b. the intensity control
   c. the vertical gain control
   d. all of the above

6. For the creation of a true Lissajous figure, the applied signals must be

   a. squarewaves of opposite polarity
   b. sawtooth waves of identical phase relationship
   c. sinewaves of equal magnitude
   d. sinewaves of unequal magnitude

7. What frequency ratio does a vertical figure 8 represent?

8. What frequency ratio limits the interpretation of a Lissajous pattern?

9. What is the difference between fluorescence, phosphorescence and luminescence?

10. Determine the unknown frequency as denoted by Figure 105E if the oscillator dial reads 5kc.

11. The primary source of electrons in the CRT is the

    a. aquadag coating
    b. second anode
    c. entire cathode surface
    d. end of the cathode facing the screen

12. Lissajous patterns are

    a. used to measure voltages
    b. patterns traced by an electron beam on the aquadag coating
    c. used to determine frequency ratios
    d. independent of the frequency but not of the phase of signals on the deflection plates

DOCID: 3726612

13. If a photograph is overexposed, should the aperture index (f-stop) be increased or decreased?

14. Why should the probe be calibrated if only sinewaves are to be observed?

15. The time/cm control is set at 1 millisecond and the multiplier is set on 2. Four cycles of the waveform are displayed across the 10 cm graticule. What is the frequency of the wave?

16. What is the function of each of the following?

    a. Astigmatism Control

    b. Time/cm

    c. AC/DC switch

    d. Triggering Mode switch

17. A certain waveform is known to have a peak-to-peak amplitude of 80 volts. The v/cm switch is set at 20 volts/cm and calibrated. The waveform observed on the oscilloscope is a trace that has a peak-to-peak deflection of only .4 cm instead of the expected 4 cm. The probable cause is

    a. improper setting of the v/cm switch
    b. the equipment is malfunctioning
    c. the 5X Magnifier is turned on
    d. probe attenuation was forgotten

18. A certain waveform is known to have a period of 250 microseconds. The wavetrain is observed on the oscilloscope with the T/cm control set at 1 millisecond and the multiplier at X2. Instead of the expected 8 cycles being displayed a little less than 1-3/4 cycles appear. The probable cause is

    a. improper setting of the Volts/cm switch
    b. the equipment is malfunctioning
    c. the 5X Magnifier is turned on
    d. the probe attenuation is too great

19. An oscillogram of a signal shows a sinewave of 15 cycles occurring in 2 cm. The recorded sweep time is 10 ms/cm. What is the frequency?

20. To correct for an underexposed oscillogram made at f/8 the aperture index should be

    a. reset to f/11
    b. reset to f/5.6
    c. unchanged
    d. none of the above

21. On the 545 oscilloscope with a sweep of 10 ms/cm, how much time is required to complete one full sweep?

    a. 1/10 second
    b. 1/2 second
    c. 1/100 second
    d. 1/25 second

DOCID: 3726612

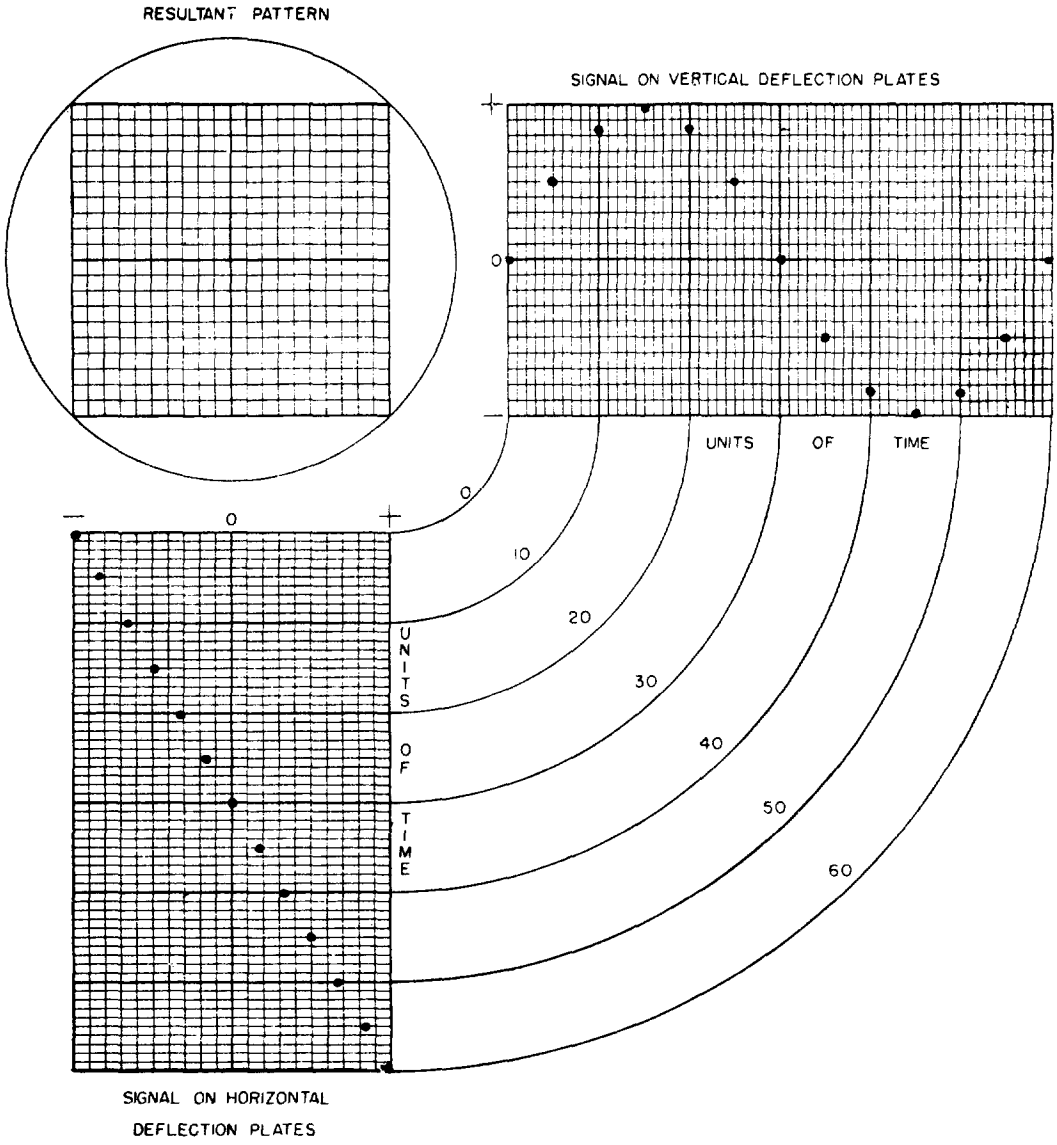


FIGURE 109

22. Plot the resulting pattern. What would be the result if the period of the sawtooth time-base and the period of the sinewave were not the same?

DOCID: 3726612

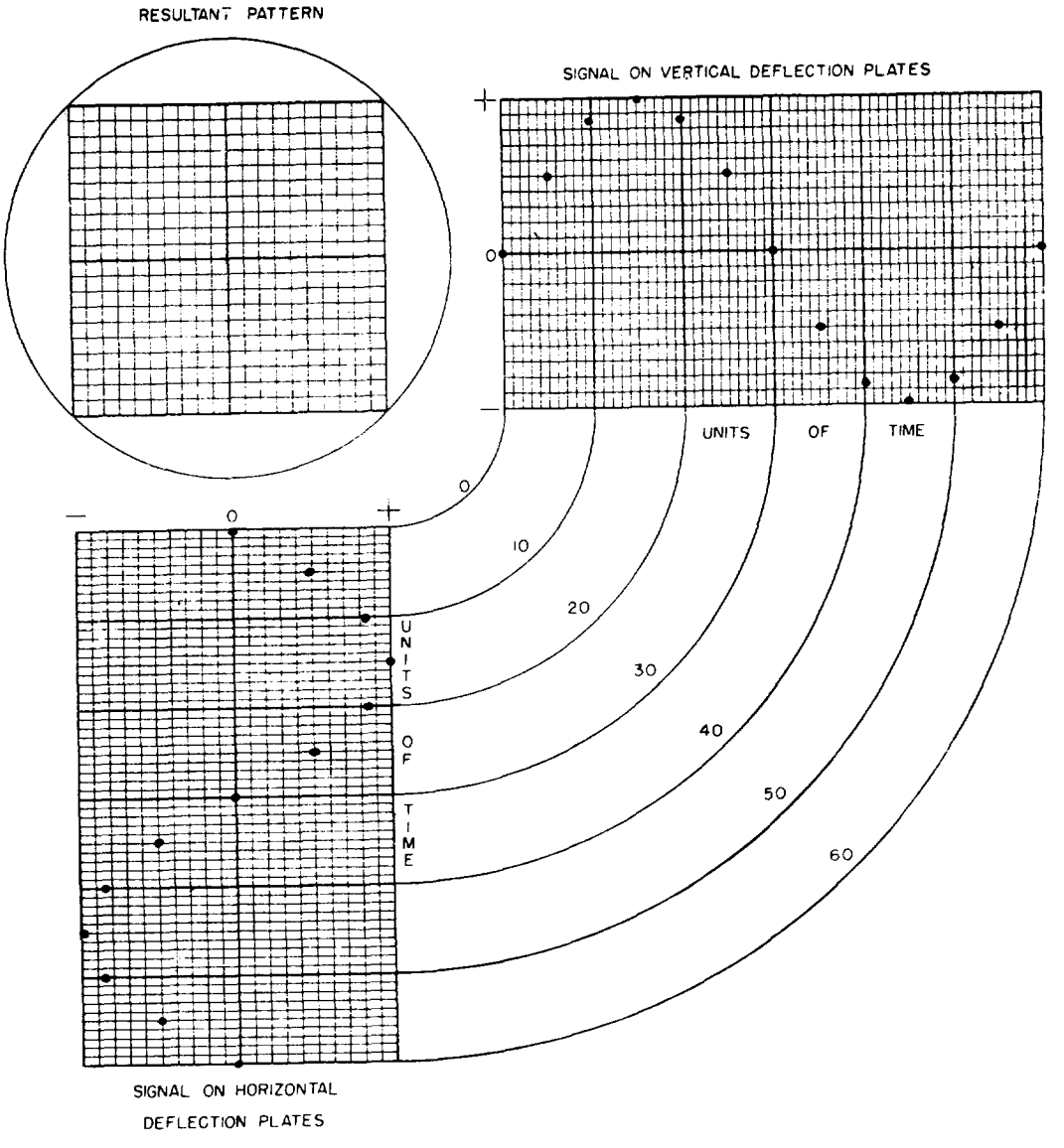


FIGURE 110.

23. Plot the resulting pattern. Find the frequency ratio and the relative phase angle. If the reference frequency is 1.5 kc what is the vertical frequency?

DOCID: 3726612

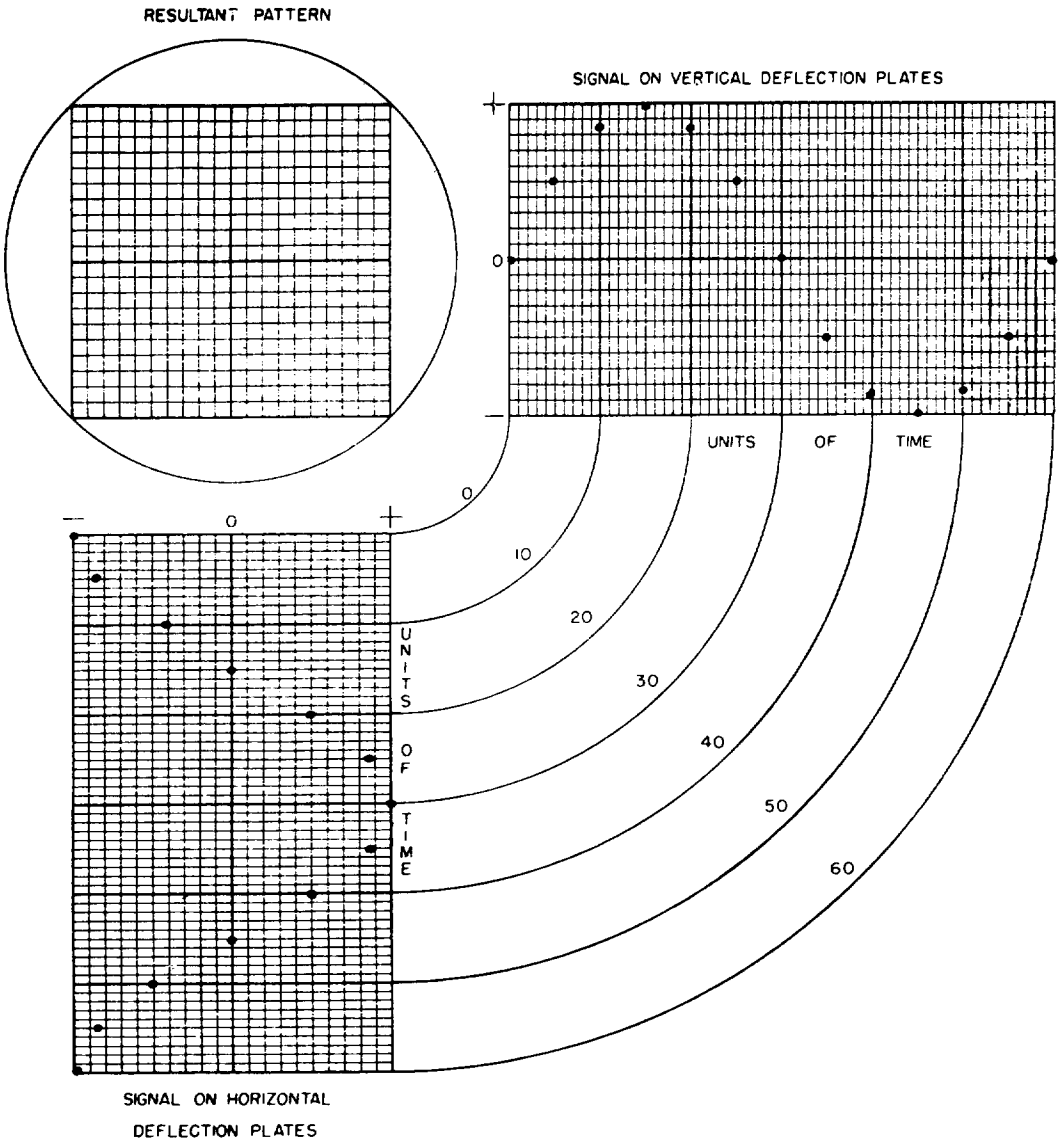


FIGURE 111.

24. **Plot the resulting pattern. Find the frequency ratio and the relative phase angle. If the reference frequency is 1.5 kc what is the vertical frequency?**

DOCID: 3726612

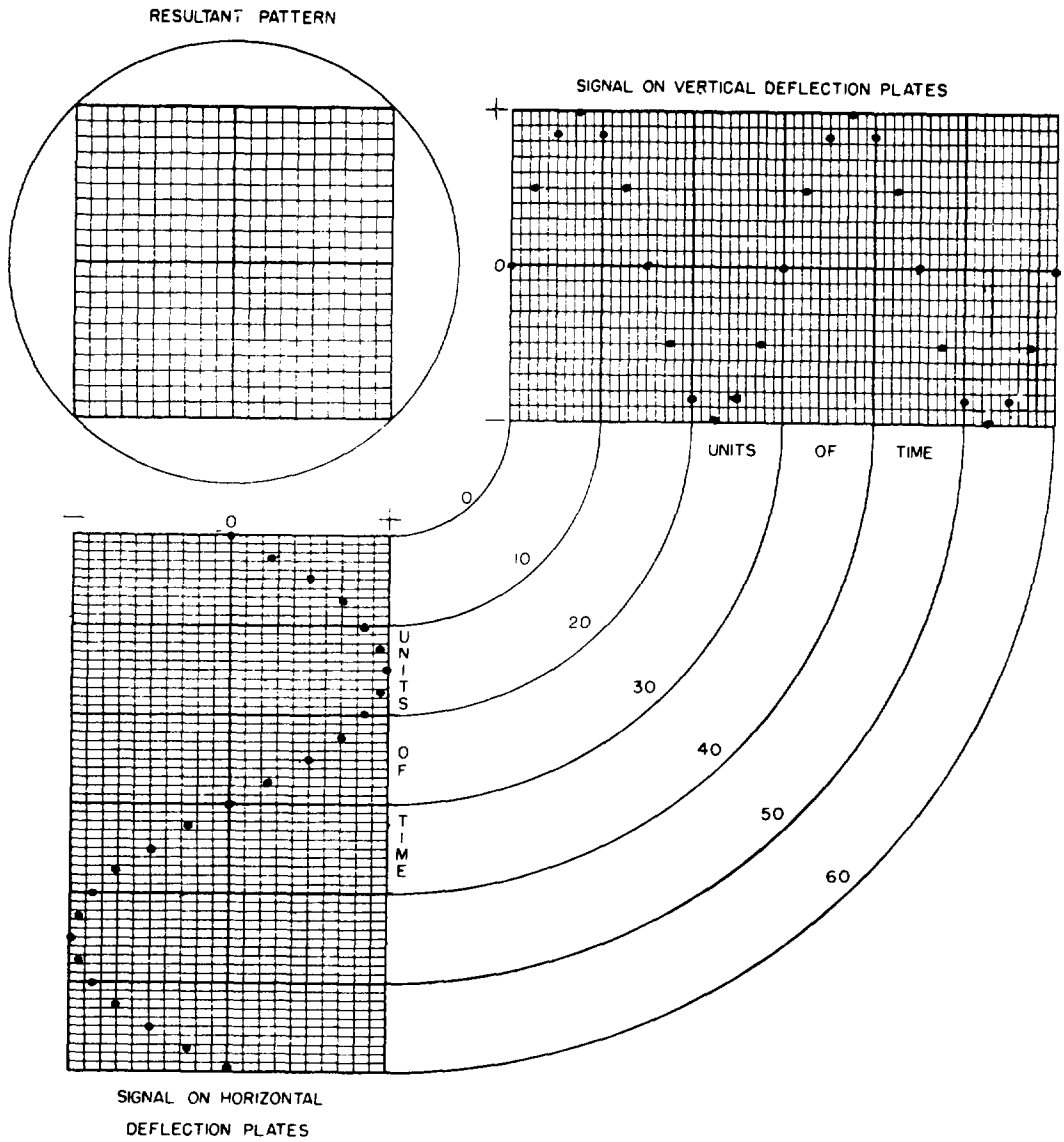


FIGURE 112.

25. Plot the resulting pattern. Find the frequency ratio. If the reference frequency is 525 cps, what is the vertical frequency?

DOCID: 3726612